

# Improved Smoothed Analysis of Multiobjective Optimization

Tobias Brunsch

Department of Computer Science  
University of Bonn, Germany  
brunsch@cs.uni-bonn.de

Heiko Röglin

Department of Computer Science  
University of Bonn, Germany  
heiko@roeglin.org

November 8, 2011

## Abstract

We present several new results about smoothed analysis of multiobjective optimization problems. Motivated by the discrepancy between worst-case analysis and practical experience, this line of research has gained a lot of attention in the last decade. We consider problems in which  $d$  linear and one arbitrary objective function are to be optimized over a set  $\mathcal{S} \subseteq \{0, 1\}^n$  of feasible solutions. We improve the previously best known bound for the smoothed number of Pareto-optimal solutions to  $O(n^{2d}\phi^d)$ , where  $\phi$  denotes the perturbation parameter. Additionally, we show that for any constant  $c$  the  $c$ -th moment of the smoothed number of Pareto-optimal solutions is bounded by  $O((n^{2d}\phi^d)^c)$ . This improves the previously best known bounds significantly. Furthermore, we address the criticism that the perturbations in smoothed analysis destroy the zero-structure of problems by showing that the smoothed number of Pareto-optimal solutions remains polynomially bounded even for zero-preserving perturbations. This broadens the class of problems captured by smoothed analysis and it has consequences for non-linear objective functions. One corollary of our result is that the smoothed number of Pareto-optimal solutions is polynomially bounded for polynomial objective functions.

## 1 Introduction

In most real-life decision-making problems there is more than one objective to be optimized. For example, when booking a train ticket, one wishes to minimize the travel time, the fare, and the number of train changes. As different objectives are often conflicting, usually no solution is simultaneously optimal in all criteria and one has to make a trade-off between different objectives. The most common way to filter out unreasonable trade-offs and to reduce the number of solutions the decision maker has to choose from is to determine the set of *Pareto-optimal solutions*, where a solution is called Pareto-optimal if no other solution is simultaneously better in all criteria.

Multiobjective optimization problems have been studied extensively in operations research and theoretical computer science (see, e.g., [9] for a comprehensive survey). In particular, many algorithms for generating the set of Pareto-optimal solutions for various optimization problems such as the (bounded) knapsack problem [16, 12], the multiobjective shortest path problem [7, 11, 19], and the multiobjective network flow problem [8, 15] have been proposed. Enumerating the set of Pareto-optimal solutions is not only used as a preprocessing step to eliminate unreasonable trade-offs, but often it is also used as an intermediate step in algorithms for solving optimization problems. For example, the Nemhauser–Ullmann algorithm [16] treats the single-criterion knapsack problem as a bicriteria optimization problem in which a solution with small weight and large profit is sought, and it generates the set of Pareto-optimal solutions, ignoring the given capacity of the knapsack. After this set has been generated, the algorithm returns the solution with the highest profit among all Pareto-optimal solutions with weight not exceeding the knapsack capacity. This solution is optimal for the given instance of the knapsack problem.

Generating the set of Pareto-optimal solutions (a.k.a. the *Pareto set*) makes only sense if few solutions are Pareto-optimal. Otherwise, it is too costly and it does not provide enough guidance to the decision maker. While, in many applications, it has been observed that the Pareto set is indeed usually small (see, e.g., [14] for an experimental study of the multiobjective shortest path problem), one can, for almost every problem with more than one objective function, easily find instances with an exponential number of Pareto-optimal solutions (see, e.g., [9]).

Motivated by the discrepancy between worst-case analysis and practical observations, *smoothed analysis* of multiobjective optimization problems has gained a lot of attention in the last decade. Smoothed analysis is a framework for judging the performance of algorithms that has been proposed in 2001 by Spielman and Teng [20] in order to explain why the simplex algorithm is efficient in practice even though it has an exponential worst-case running time. In this framework, inputs are generated in two steps: first, an adversary chooses an arbitrary instance, and then this instance is slightly perturbed at random. The smoothed performance of an algorithm is defined to be the worst expected performance the adversary can achieve. This model can be viewed as a less pessimistic worst-case analysis, in which the randomness rules out pathological worst-case instances that are rarely observed in practice but dominate the worst-case analysis. If the smoothed running time of an algorithm is low and inputs are subject to a small amount of random noise, then it is unlikely to encounter an instance on which the algorithm performs poorly. In practice, random noise can stem from measurement errors, numerical imprecision or rounding errors. It can also model arbitrary influences, which we cannot quantify exactly, but for which there is also no reason to believe that they are adversarial.

After its invention in 2001, smoothed analysis has been successfully applied in a variety of contexts, e.g., to explain the practical success of local search methods, heuristics for the knapsack problem, online algorithms, and clustering. A recent survey by Spielman and Teng [21] summarizes some of these results. One of the areas in which smoothed analysis has been applied extensively is multiobjective optimization. In 2003 Beier and Vöcking [3] initiated this line of research by showing that the smoothed number of Pareto-optimal solutions is polynomially bounded for any linear binary optimization problem with two objective functions. This was the first rigorous explanation why heuristics for generating the set of Pareto-optimal solutions are successful in practice despite their bad worst-case behavior. In the last years, Beier and Vöcking’s original result has been improved and extended significantly in a series of papers. A discussion of this work follows in the next section after the formal description of the model.

## 1.1 Model and Previous Work

We consider a very general model of multiobjective optimization problems. An instance of such a problem consists of  $d + 1$  objective functions  $V^1, \dots, V^{d+1}$  that are to be optimized over a set  $\mathcal{S} \subseteq \{0, 1\}^n$  of feasible solutions. While the set  $\mathcal{S}$  and the last objective function  $V^{d+1}: \mathcal{S} \rightarrow \mathbb{R}$  can be arbitrary, the first  $d$  objective functions have to be linear of the form  $V^t(x) = V_1^t x_1 + \dots + V_n^t x_n$  for  $x = (x_1, \dots, x_n) \in \mathcal{S}$  and  $t \in \{1, \dots, d\}$ . We assume without loss of generality that all objectives are to be minimized and we call a solution  $x \in \mathcal{S}$  *Pareto-optimal* if there does not exist a solution  $y \in \mathcal{S}$  with  $V^t(y) \leq V^t(x)$  for all  $t \in \{1, \dots, d + 1\}$  and  $V^t(y) < V^t(x)$  for at least one  $t \in \{1, \dots, d + 1\}$ .

If one is allowed to choose the set  $\mathcal{S}$ , the objective function  $V^{d+1}$ , and the coefficients of the linear objective functions arbitrarily, then, even for  $d = 1$ , one can easily construct instances with an exponential number of Pareto-optimal solutions. For this reason Beier and Vöcking introduced the model of  $\phi$ -smooth instances [3], in which an adversary can choose the set  $\mathcal{S}$  and the objective function  $V^{d+1}$  arbitrarily while he can only specify a probability density function  $f_i^t: [-1, 1] \rightarrow [0, \phi]$  for each coefficient  $V_i^t$  according to which it is chosen independently from the other coefficients. This model is more general than Spielman and Teng’s original two-step model in which the adversary first chooses coefficients which are afterwards subject to Gaussian perturbations. In  $\phi$ -smooth

instances the adversary can additionally determine the type of noise. He could, for example, specify for each coefficient an interval of length  $1/\phi$  from which it is chosen uniformly at random. The parameter  $\phi \geq 1$  can be seen as a measure for the power of the adversary: the larger  $\phi$  the more precisely he can specify the coefficients of the linear objective functions. The aforementioned example of uniform distributions shows that for  $\phi \rightarrow \infty$  smoothed analysis becomes a worst-case analysis.

The *smoothed number of Pareto-optimal solutions* depends on the number  $n$  of binary variables and the perturbation parameter  $\phi$ . It is defined to be the largest expected number of Pareto-optimal solutions the adversary can achieve by any choice of  $\mathcal{S} \subseteq \{0, 1\}^n$ ,  $V^{d+1}: \mathcal{S} \rightarrow \mathbb{R}$ , and the densities  $f_i^t: [-1, 1] \rightarrow [0, \phi]$ . In the following we assume that the adversary has made arbitrary fixed choices for these entities. Then we can associate with every matrix  $V \in \mathbb{R}^{d \times n}$  the number  $\text{PO}(V)$  of Pareto-optimal solutions in  $\mathcal{S}$  when the coefficients  $V_i^t$  of the  $d$  linear objective functions take the values given in  $V$ . Assuming that the adversary has made worst-case choices for  $\mathcal{S}$ ,  $V^{d+1}$ , and the densities  $f_i^t$ , the *smoothed number of Pareto-optimal solutions* is the expected value  $\mathbf{E}_V[\text{PO}(V)]$ , where the coefficients in  $V$  are chosen according to the densities  $f_i^t$ . For  $c \geq 1$ , we call  $\mathbf{E}_V[\text{PO}^c(V)]$  the *c-th moment of the smoothed number of Pareto-optimal solutions*.

Beier and Vöcking [3] showed that for the bicriteria case (i.e.,  $d = 1$ ) the smoothed number of Pareto-optimal solutions is  $O(n^4\phi)$  and  $\Omega(n^2)$ . The upper bound was later simplified and improved by Beier et al. [2] to  $O(n^2\phi)$ . In his PhD thesis [1], Beier conjectured that the smoothed number of Pareto-optimal solutions is polynomially bounded in  $n$  and  $\phi$  for every constant  $d$ . This conjecture was proven by Röglin and Teng [17], who showed that for any fixed  $d \geq 1$ , the smoothed number of Pareto-optimal solutions is  $O((n^2\phi)^{f(d)})$ , where the function  $f$  is roughly  $f(d) = 2^d d!$ . They also proved that for any constant  $c$  the  $c$ -th moment of the smoothed number of Pareto-optimal solutions is bounded by  $O((n^2\phi)^{c \cdot f(d)})$ . Recently, Moitra and O'Donnell [13] improved the bound for the smoothed number of Pareto-optimal solutions significantly to  $O(n^{2d}\phi^{d(d+1)/2})$ . The bound for the moments could, however, not be improved by their methods. Very recently a combination of results by Goyal and Rademacher [10] and our results in [6] was observed to yield a lower bound of  $\Omega(n^{d-1.5}\phi^d)$  [10].

## 1.2 Our Results

In this paper, we present several new results about smoothed analysis of multiobjective binary optimization problems. Besides general  $\phi$ -smooth instances, we consider for some results only the special case of *quasiconcave density functions*. This means that we assume that every coefficient  $V_i^t$  is chosen independently according to its own density function  $f_i^t: [-1, 1] \rightarrow [0, \phi]$  with the additional requirement that for every density  $f_i^t$  there is a value  $x_i^t \in [-1, 1]$  such that  $f_i^t$  is non-decreasing in the interval  $[-1, x_i^t]$  and non-increasing in the interval  $[x_i^t, 1]$ . We do not think that this is a severe restriction because all natural perturbation models, like Gaussian or uniform perturbations, use quasiconcave density functions. Furthermore, quasiconcave densities capture the essence of a perturbation: each coefficient  $V_i^t$  has an unperturbed value  $x_i^t$  and the probability that the perturbed coefficient takes a value  $z$  becomes smaller with increasing distance  $|z - x_i^t|$ . We will call these instances *quasiconcave  $\phi$ -smooth instances* in the following.

Beier and Vöcking originally only considered  $\phi$ -smooth instances for bicriteria optimization problems (i.e., for the case  $d = 1$ ). The above described canonical generalization of this model to multiobjective optimization problems, on which Röglin and Teng's [17] and Moitra and O'Donnell's results [13] are based, appears to be very general and flexible on the first glance. However, one aspect limits its applicability severely and makes it impossible to formulate certain multiobjective linear optimization problems in this model. The weak point of the model is that it assumes that every binary variable  $x_i$  appears in every linear objective function as it is not possible to set some coefficients  $V_i^t$  deterministically to zero.

Already Spielman and Teng [20] and Beier and Vöcking [4] observed that the zeros often encode

an essential part of the combinatorial structure of a problem and they suggested to analyze *zero-preserving perturbations* in which it is possible to either choose a density  $f_i^t$  according to which the coefficient  $V_i^t$  is chosen or to set it deterministically to zero. Zero-preserving perturbations have been studied in [18] and [4] for analyzing smoothed condition numbers of matrices and the smoothed complexity of binary optimization problems. For the smoothed number of Pareto-optimal solutions no upper bounds are known that are valid for zero-preserving perturbations (except trivial worst-case bounds), and in particular the bounds proven in [17] and [13] do not seem to generalize easily to zero-preserving perturbations. In this paper, we develop new techniques for analyzing the smoothed number of Pareto-optimal solutions that can also be used for analyzing zero-preserving perturbations.

**Theorem 1.** *For any  $d \geq 1$ , the smoothed number of Pareto-optimal solutions is  $O(n^{d^3+d^2+d}\phi^d)$  for quasiconcave  $\phi$ -smooth instances with zero-preserving perturbations and  $O((n\phi)^{d^3+d^2+d})$  for general  $\phi$ -smooth instances with zero-preserving perturbations.*

In Section 1.3 we will present some applications of zero-preserving perturbations. We will see that they allow us not only to extend the smoothed analysis to linear multiobjective optimization problems that are not captured by the previous model without zero-preserving perturbations, but that they also enable us to bound the smoothed number of Pareto-optimal solutions in problems with *non-linear objective functions*. In particular, the number of Pareto-optimal solutions for polynomial objective functions can be bounded by Theorem 1. We say that a  $\phi$ -smooth instance has polynomial objective functions if every objective function  $V^t$ ,  $t \in \{1, \dots, d\}$ , is the weighted sum of at most  $m$  monomials, where the adversary can specify a  $\phi$ -bounded density on  $[0, 1]$  for every weight according to which it is chosen.

**Corollary 2.** *For any  $d \geq 1$ , the smoothed number of Pareto-optimal solutions is  $O(m^{d^3+d^2+d}\phi^d)$  for quasiconcave  $\phi$ -smooth instances with zero-preserving perturbations and polynomial objective functions. For general  $\phi$ -smooth instances with zero-preserving perturbations and polynomial objective functions the smoothed number of Pareto-optimal solutions is  $O((m\phi)^{d^3+d^2+d})$ .*

In addition to zero-preserving perturbations we also study the standard model of  $\phi$ -smooth instances. We present significantly improved bounds for the smoothed number of Pareto-optimal solutions and the moments, answering two questions posed by Moitra and O'Donnell [13].

**Theorem 3.** *For any  $d \geq 1$ , the smoothed number of Pareto-optimal solutions is  $O(n^{2d}\phi^d)$  for quasiconcave  $\phi$ -smooth instances.*

This improves the previously best known bound of  $O(n^{2d}\phi^{d(d+1)/2})$  (which is, however, valid also for non-quasiconcave densities) and it answers a question posed by Moitra and O'Donnell whether it is possible to improve the factor of  $\phi^{d(d+1)/2}$  in their bound [13]. Together with the recent lower bound of  $\Omega(n^{d-1.5}\phi^d)$  [10], which is also valid for quasiconcave density functions, this shows that the exponents of both  $n$  and  $\phi$  are linear in  $d$ .

**Theorem 4.** *For any  $d \geq 1$  and any constant  $c \in \mathbb{N}$ , the  $c$ -th moment of the smoothed number of Pareto-optimal solutions is  $O((n^{2d}\phi^d)^c)$  for quasiconcave  $\phi$ -smooth instances and  $O((n^{2d}\phi^{d(d+1)})^c)$  for general  $\phi$ -smooth instances.*

This answers a question in [13] whether it is possible to improve the bounds for the moments in [17] and it yields better concentration bounds for the smoothed number of Pareto-optimal solutions. Our results also have immediate consequences for the expected running times of various algorithms because most heuristics for generating the Pareto set of some problem (including the ones mentioned at the beginning of the introduction) have a running time that depends linearly or quadratically on the size of the Pareto set. The improved bounds on the smoothed number of Pareto-optimal solutions and the second moment of this number yield improved bounds on the smoothed running times of these algorithms.

### 1.3 Applications of Zero-preserving Perturbations

Let us first of all remark that in the bicriteria case, which was studied in [3], zero-preserving perturbations are not more powerful than other perturbations because they can be simulated by the right choice of  $\mathcal{S} \subseteq \{0,1\}^n$  and the objective function  $V^2: \mathcal{S} \rightarrow \mathbb{R}$ . Assume, for example, that the adversary has chosen  $\mathcal{S}$  and  $V^2$  and has decided that the first coefficient  $V_1^1$  of the first objective function should be deterministically set to zero. Also assume without loss of generality that  $V^2$  assigns unique values to the solutions in  $\mathcal{S}$ . If this is not the case, we can make  $V^2$  injective without reducing the smoothed number of Pareto-optimal solutions. We can partition the set  $\mathcal{S}$  in classes of solutions that agree in all components except for the first one. That is, two solutions  $x \in \mathcal{S}$  and  $y \in \mathcal{S}$  belong to the same class if  $x_i = y_i$  for all  $i \in \{2, \dots, n\}$ . All solutions in the same class have the same value in the first objective  $V^1$  as they differ only in the binary variable  $x_1$ , whose coefficient has been set to zero. We construct a new set of solutions  $\mathcal{S}'$  that contains for every class only the solution with smallest value in  $V^2$ . It is easy to see that the number of Pareto-optimal solutions is the same with respect to  $\mathcal{S}$  and with respect to  $\mathcal{S}'$  because all solutions in  $\mathcal{S} \setminus \mathcal{S}'$  are dominated by solutions in  $\mathcal{S}'$ .

Then we transform the set  $\mathcal{S}' \subseteq \{0,1\}^n$  into a set  $\mathcal{S}'' \subseteq \{0,1\}^{n-1}$  by dropping the first component of every solution. Furthermore, we define a function  $W^2: \mathcal{S}'' \rightarrow \mathbb{R}$  that assigns to every solution  $x \in \mathcal{S}''$  the same value that  $V^2$  assigns to the corresponding solution in  $\mathcal{S}'$ . One can easily verify that the Pareto set with respect to  $\mathcal{S}'$  and  $V^2$  is identical with the Pareto-set with respect to  $\mathcal{S}''$  and  $W^2$ . The only difference is that in the latter problem we have eliminated the coefficient that is deterministically set to zero. Such an easy reduction of zero-preserving perturbations to other perturbations does not seem to be possible for  $d \geq 2$  anymore.

#### Path Trading

Berger et al. [5] study a model for routing in networks. In their model there is a graph  $G = (V, E)$  whose vertex set  $V$  is partitioned into mutually disjoint sets  $V_1, \dots, V_k$ . We can think of  $G$  as the Internet graph whose vertices are owned and controlled by  $k$  different autonomous systems (ASes). We denote by  $E_i \subseteq E$  the set of edges inside  $V_i$ . The graph  $G$  is undirected, and each edge  $e \in E$  has a length  $\ell_e \in \mathbb{R}_{\geq 0}$ . The traffic is modeled by a set of requests, where each request is characterized by its source node  $s \in V$  and its target node  $t \in V$ . We consider the case that a single request has to be routed from a source  $s \in V_1$  to a target  $t \in V_k$ . The Border Gateway Protocol (BGP) determines for this request the order in which it has to be routed through the ASes. We assume without loss of generality that the request has to be routed through  $V_1, V_2, \dots, V_k$  in this order, and we say that a path  $P$  is valid if it connects  $s$  to  $t$  and visits the ASes in the order specified by the BGP protocol. This means that the first AS has to choose a path  $P_1$  inside  $V_1$  from  $s$  to some node in  $V_1$  that is connected to some node  $v_2 \in V_2$ . Then the second AS has to choose a path  $P_2$  inside  $V_2$  from  $v_2$  to some node in  $V_2$  that is connected to some node  $v_3 \in V_3$  and so on. For simplicity, the costs of routing a packet between two ASes are assumed to be zero, whereas AS  $i$  incurs costs of  $\sum_{e \in P_i} \ell_e$  for routing the packet inside  $V_i$  along path  $P_i$ . In the common hot-potato routing, every AS is only interested in minimizing its own costs. To model this, there are  $k$  objective functions that map each valid path  $P$  to a cost vector  $(C_1(P), \dots, C_k(P))$ , where

$$C_i(P) = \sum_{e \in P \cap E_i} \ell_e \quad \text{for } i \in \{1, \dots, k\}.$$

In [5] the problem of *path trading* is considered. This problem considers the question whether ASes can reduce their costs if they deviate from the hot-potato strategy and coordinate the chosen paths for multiple requests. A main ingredient in the algorithm proposed in [5] to solve this problem is to generate the set of Pareto-optimal valid paths with respect to the  $k$  objective functions defined above. As this Pareto set can be exponentially large in the worst case, Berger et al. proposed to study  $\phi$ -smooth instances in which an adversary chooses the graph  $G$  and a density  $f_e: [0,1] \rightarrow$

$[0, \phi]$  for every edge length  $\ell_e$  according to which it is chosen. It seems as if we could easily apply the results in [17] and [13] to bound the smoothed number of Pareto-optimal paths. If we set  $S \subseteq \{0, 1\}^{|E|}$  to be the set of incidence vectors of valid paths, then all objective functions  $C_i$  are linear in the binary variables  $x_e$ ,  $e \in E$ . For  $x \in S$ , we have

$$C_i(x) = \sum_{e \in E_i} \ell_e x_e.$$

Note, however, that different objective functions contain different variables  $x_e$  because the coefficients of all  $x_e$  with  $e \notin E_i$  are set to zero in  $C_i$ . This is an important combinatorial property of the path trading problem that has to be obeyed. In the model in [17] and [13] it is not possible to set coefficients deterministically to zero. The best we can do is to replace each zero by a uniform density on the interval  $[0, 1/\phi]$ . Then, however, an AS would incur positive costs for any edge that is used and not only for its own edges, which does not resemble the structure of the problem. Theorem 1, which allows zero-preserving perturbations, yields immediately the following result.

**Corollary 5.** *The smoothed number of Pareto-optimal valid paths is polynomially bounded in  $|E|$  and  $\phi$  for any constant  $k$ .*

### Non-linear Objective Functions

Even though we assumed above that the objective functions  $V^1, \dots, V^d$  are linear, we can also extend the smoothed analysis to non-linear objective functions. We consider first the bicriteria case  $d = 1$ . As above, we assume that the adversary has chosen a set  $\mathcal{S} \subseteq \{0, 1\}^n$  of feasible solutions and an arbitrary injective objective function  $V^2: \mathcal{S} \rightarrow \mathbb{R}$ . In addition to that the adversary can choose  $m_1$  arbitrary functions  $I_i^1: \mathcal{S} \rightarrow \{0, 1\}$ ,  $i \in \{1, \dots, m_1\}$ . The objective function  $V^1: \mathcal{S} \rightarrow \mathbb{R}$  is defined to be a weighted sum of the functions  $I_i^1$ :

$$V^1(x) = \sum_{i=1}^{m_1} w_i^1 I_i^1(x),$$

where each weight  $w_i^1$  is randomly chosen according to a density  $f_i^1: [0, 1] \rightarrow [0, \phi]$  chosen by the adversary. There is a wide variety of functions  $V^1(x)$  that can be expressed in this way. We can, for example, express every polynomial if we let  $I_1^1, \dots, I_{m_1}^1$  be its monomials.

We can linearize the problem by introducing a binary variable for every function  $I_i^1$ . Then the set of feasible solutions becomes

$$\mathcal{S}' = \left\{ y \in \{0, 1\}^{m_1} \mid \exists x \in \mathcal{S} : \forall i \in \{1, \dots, m_1\} : I_i^1(x) = y_i \right\}.$$

For this set of feasible solutions we define  $W^1: \mathcal{S}' \rightarrow \mathbb{R}$  and  $W^2: \mathcal{S}' \rightarrow \mathbb{R}$  as follows:

$$W^1(y) = \sum_{j=1}^{m_1} w_j y_j \quad \text{and} \quad W^2(y) = \min \left\{ V^2(x) \mid \forall i \in \{1, \dots, m_1\} : I_i^1(x) = y_i \right\}.$$

One can easily verify that the problem defined by  $\mathcal{S}$ ,  $V^1$ , and  $V^2$  and the problem defined by  $\mathcal{S}'$ ,  $W^1$ , and  $W^2$  are equivalent and have the same number of Pareto-optimal solutions. The latter problem is linear and hence we can apply the result by Beier et al. [2], which yields that the smoothed number of Pareto-optimal solutions is bounded by  $O(m_1^2 \phi)$ . This shows in particular that the smoothed number of Pareto-optimal solutions is polynomially bounded in the number of monomials and the density parameter for every polynomial objective function  $V^1$ .

We can easily extend these considerations to multiobjective problems with  $d \geq 2$ . For these problems the adversary chooses a set  $\mathcal{S} \subseteq \{0, 1\}^n$ , numbers  $m_1, \dots, m_d \in \mathbb{N}$ , and an arbitrary injective objective function  $V^{d+1}: \mathcal{S} \rightarrow \mathbb{R}$ . In addition to that he chooses arbitrary functions  $I_i^t: \mathcal{S} \rightarrow \{0, 1\}$  for  $t \in \{1, \dots, d\}$  and  $i \in \{1, \dots, m_t\}$ . Every objective function  $V^t: \mathcal{S} \rightarrow \mathbb{R}$  is

a weighted sum of the functions  $I_t^1$ :

$$V^t(x) = \sum_{i=1}^{m_t} w_i^t I_i^t(x),$$

where each weight  $w_i^t$  is randomly chosen according to a density  $f_i^t: [0, 1] \rightarrow [0, \phi]$  chosen by the adversary. Similar as the bicriteria case, also this problem can be linearized. However, the previous results about the smoothed number of Pareto-optimal solutions can only be applied if every objective function  $V^t$  is composed of exactly the same functions  $I_i^t$ . Theorem 1 implies that the smoothed number of Pareto-optimal solutions is polynomially bounded in  $\sum m_i$  and  $\phi$ , for any choice of the  $I_i^t$ .

## Outline

After introducing some notation in the next section, we present an outline of our approach and our methods in Section 3. In Section 4 we prove Theorems 3 and 4. In Section 5 we consider zero-preserving perturbations and prove Theorem 1. We conclude the paper with some open questions.

## 2 Notation

For the sake of simplicity we write  $V^t x$  instead of  $V^t(x)$ , even for the adversarial objective  $V^{d+1}$ . With  $V^{k_1 \dots k_t} x$  we refer to the vector  $(V^{k_1} x, \dots, V^{k_t} x)$ . In our analysis, we will shift the solutions  $x \in \mathcal{S}$  by a certain vector  $u \in \{0, 1\}^n$  and consider the values  $V^t \cdot (x - u)$ . For the linear objectives we mean the value  $V^t x - V^t u$ , where  $V^t u$  is well-defined even for a shift vector  $u \in \{0, 1\}^n \setminus \mathcal{S}$ . For the adversarial objective, however, we define  $V^{d+1} \cdot (x - u) := V^{d+1} x$ . It should not be confused with  $V^{d+1} y$  for  $y = x - u$ . In the whole paper let  $\varepsilon > 0$  be a small real for which  $1/\varepsilon$  is integral. Let  $b = (b_1, \dots, b_d) \in \mathbb{R}^d$  be a vector such that  $b_k$  is an integral multiple of  $\varepsilon$  for any  $k$ . We will call the set  $B = \{(y_1, \dots, y_d) \in \mathbb{R}^d : y_k \in (b_k, b_k + \varepsilon] \text{ for any } k\}$  an  $\varepsilon$ -box and  $b$  the corner of  $B$ . For a vector  $x \in \{-1, 0, 1\}^n$  the expression  $B_V(x)$  denotes the unique  $\varepsilon$ -box  $B$  for which  $V^{1 \dots d} x \in B$ . We call  $B$  the  $\varepsilon$ -box of  $x$  and say that  $x$  lies in  $B$ . With  $\mathbb{B}_\varepsilon$  we denote the set of all  $\varepsilon$ -boxes having corners  $b$  for which  $b \in \{-n/\varepsilon, \dots, n/\varepsilon - 1\}^d$ . Hence,  $|\mathbb{B}_\varepsilon| = (2n/\varepsilon)^d$ . If all coefficients  $V_i^k$  of  $V$  are from  $[-1, 1]$ , which is true for any of the models considered in this paper, and if for any  $k = 1, \dots, d$  there is an index  $i$  such that  $|V_i^k| < 1$ , which holds with probability 1 in any of our models, then the  $\varepsilon$ -box of any vector  $x \in \{-1, 0, 1\}^n$  belongs to  $\mathbb{B}_\varepsilon$ . Note, that all vectors  $x$  constructed in this paper are from  $\{-1, 0, 1\}^n$ . Hence, without any further explanation we will use that  $B_V(x) \in \mathbb{B}_\varepsilon$ .

In this paper we extensively use tuples instead of sets. The reason for this is that we are not only interested in certain components of a vector or a matrix, but we also want to describe in which order they are considered. This will be clear after the introduction of the following notation. Let  $n, m$  be positive integers and let  $a_1, \dots, a_n, b_1, \dots, b_m$  be arbitrary reals. We define  $[n] = (1, \dots, n)$ ,  $[n]_0 = (0, 1, \dots, n)$ ,  $|(a_1, \dots, a_n)| = n$  and  $(a_1, \dots, a_n) \cup (b_1, \dots, b_m) = (a_1, \dots, a_n, b_1, \dots, b_m)$ . By  $(a_1, \dots, a_n) \setminus (b_1, \dots, b_m)$  and  $(a_1, \dots, a_n) \cap (b_1, \dots, b_m)$  we denote the tuples we obtain when we remove all elements from  $(a_1, \dots, a_n)$  that (do not) belong to  $(b_1, \dots, b_m)$ . We write  $(a_1, \dots, a_n) \subseteq (b_1, \dots, b_m)$  if  $a_k \in (b_1, \dots, b_m)$  for any index  $k \in [n]$ . Let  $x$  be a vector and let  $A$  be a matrix. By  $x|_{i_1 \dots i_n} = x|_{(i_1, \dots, i_n)}$  we denote the column vector  $(x_{i_1}, \dots, x_{i_n})$ , by  $A|_{(i_1, \dots, i_n)}$  we denote the matrix consisting of the rows  $i_1, \dots, i_n$  of matrix  $A$  in this order.

For an index set  $I \subseteq [n]$  and a vector  $y \in \{0, 1\}^n$  let  $\mathcal{S}_I(y)$  denote the set of all solutions  $z \in \mathcal{S}$  such that  $z_i = y_i$  for any index  $i \in I$ . For the sake of simplicity we also use the notation  $\mathcal{S}_I(\hat{y})$  to describe the set  $\{z \in \mathcal{S} : z_i = \hat{y}_i\}$  for a vector  $\hat{y} \in \{0, 1\}^{|I|}$  when the components of  $y$  are labeled by  $y_{i_1}, \dots, y_{i_{|I|}}$  where  $I = (i_1, \dots, i_{|I|})$ .

With  $\mathbb{I}_n$  we refer to the  $n \times n$ -diagonal matrix  $\text{diag}(1, \dots, 1)$  and with  $\mathbb{O}_{m \times n}$  to the  $m \times n$ -matrix whose entries are all zero. If the number of rows and columns are clear, then we drop the indices.

### 3 Outline of our Approach

To prove our results we adapt and improve methods from the previous analyses by Moitra and O'Donnell [13] and by Röglin and Teng [17] and combine them in a novel way. Since all coefficients of the linear objective functions lie in the interval  $[-1, 1]$ , for every solution  $x \in \mathcal{S}$  the vector  $V^{1 \dots d}x$  lies in the hypercube  $[-n, n]^d$ . The first step is to partition this hypercube into  $\varepsilon$ -boxes. If  $\varepsilon$  is very small (exponentially small in  $n$ ), then it is unlikely that there are two different solutions  $x \in \mathcal{S}$  and  $y \in \mathcal{S}$  that lie in the same  $\varepsilon$ -box  $B$  unless  $x$  and  $y$  differ only in positions that are not perturbed in any of the objective functions, in which case we consider them as the same solution. In the remainder of this section we assume that no two solutions lie in the same  $\varepsilon$ -box. Then, in order to bound the number of Pareto-optimal solutions, it suffices to count the number of non-empty  $\varepsilon$ -boxes.

In order to prove Theorem 3 we show that for any fixed  $\varepsilon$ -box the probability that it contains a Pareto-optimal solution is bounded by  $\kappa n^d \phi^d \varepsilon^d$  for a sufficiently large constant  $\kappa$ . This implies the theorem as the number of  $\varepsilon$ -boxes is  $(2n/\varepsilon)^d$ . Fix an arbitrary  $\varepsilon$ -box  $B$ . In the following we will call a solution  $x \in \mathcal{S}$  a *candidate* if there is a realization of  $V$  such that  $x$  is Pareto-optimal and lies in  $B$ . If there was only a single candidate  $x \in \mathcal{S}$ , then we could bound the probability that there is a Pareto-optimal solution in  $B$  by the probability that this particular solution  $x$  lies in  $B$ . This probability can easily be bounded from above by  $\varepsilon^d \phi^d$ . However, in principle, every solution  $x \in \mathcal{S}$  can be a candidate and a union bound over all of them leads to a factor of  $|\mathcal{S}|$  in the bound, which we have to avoid.

Following ideas of Moitra and O'Donnell, we divide the draw of the random matrix  $V$  into two steps. In the first step some information about  $V$  is revealed that suffices to limit the set of candidates to a single solution  $x \in \mathcal{S}$ . The exact position  $V^{1 \dots d}x$  of this solution is determined in the second step. If the information that is revealed in the two steps is chosen carefully, then there is enough randomness left in the second step to bound the probability that  $x$  lies in the  $\varepsilon$ -box  $B$ . In Moitra and O'Donnell's analysis the coefficients in the matrix  $V$  are partitioned into two groups. In the first step the first group of coefficients is drawn, which suffices to determine the unique candidate  $x$ , and in the second step the remaining coefficients are drawn, which suffices to bound the probability that  $x$  lies in  $B$ . The second part consists essentially of  $d(d+1)/2$  coefficients, which causes the factor of  $\phi^{d(d+1)/2}$  in their bound.

We improve the analysis by a different choice of how to break the draw of  $V$  into two parts. As in the previous analysis, most coefficients are drawn in the first step. Only  $d^2$  coefficients of  $V$  are drawn in the second step. However, these coefficients are not left completely random as in [13] because after the other coefficients have been drawn there can still be multiple candidates for Pareto-optimal solutions in  $B$ . Instead, the randomness is reduced further by drawing  $d(d-1)$  linear combinations of these random variables in the first step. These linear combinations have the property that, after they have been drawn, there is a unique candidate  $x$  whose position can be described by  $d$  linear combinations that are linearly independent of the linear combinations already drawn in the first step. In [17] it was observed that linearly independent linear combinations of independent random variables behave in some respect similar to independent random variables. With this insight one can argue that in the second step there is still enough randomness to bound the probability that  $x$  lies in  $B$ . While the analysis in [17] yields only a bound proportional to  $\phi^{d^2} \varepsilon^d$ , we prove an improved result for quasiconcave densities that yields the desired bound proportional to  $\phi^d \varepsilon^d$ .

For analyzing higher moments, it does not suffice to bound the probability that a fixed  $\varepsilon$ -box contains a Pareto-optimal solution. Instead, in order to bound the  $c^{\text{th}}$  moment, we sum over all  $c$ -tuples  $(B_1, \dots, B_c)$  of  $\varepsilon$ -boxes the probability that all  $\varepsilon$ -boxes  $B_1, \dots, B_c$  contain simultaneously a Pareto-optimal solution. We bound this probability from above by  $\kappa n^{cd} \phi^{cd} \varepsilon^{cd}$  for a sufficiently large constant  $\kappa$ . Since there are  $(2n/\varepsilon)^{cd}$  different  $c$ -tuples of  $\varepsilon$ -boxes, this implies the bound of  $O((n^2 \phi)^{cd})$  for the smoothed  $c^{\text{th}}$  moment of the number of Pareto-optimal solutions.



Let us fix a  $c$ -tuple  $(B_1, \dots, B_c)$  of  $\varepsilon$ -boxes. The approach to bound the probability that all of these  $\varepsilon$ -boxes contain simultaneously a Pareto-optimal solution is similar to the approach for the first moment. We divide the draw of  $V$  into two steps. In the first step enough information is revealed to identify for each of the  $\varepsilon$ -boxes  $B_i$  a unique candidate  $x_i \in \mathcal{S}$  for a Pareto-optimal solution in  $B_i$ . If we do this carefully, then there is enough randomness left in the second step to bound the probability that  $V^{1\dots d}x_i \in B_i$  for every  $i \in [c]$ . Again most coefficients are drawn in the first step and some linear combinations of the other  $cd^2$  coefficients are also drawn in the first step. However, we cannot simply repeat the construction for the first moment independently  $c$  times because then there might be dependencies between the events  $V^{1\dots d}x_i \in B_i$  for different  $i$ . In order to bound the probability that in the second step all  $x_i$  lie in their corresponding  $\varepsilon$ -boxes  $B_i$ , we need to ensure that the events  $V^{1\dots d}x_i \in B_i$  are (almost) independent after the information from the first step has been revealed.

The general approach to handle zero-preserving perturbations is closely related to the approach for bounding the first moment for non-zero-preserving perturbations. However, additional complications have to be handled. The main problem is that we cannot easily guarantee anymore that the linear combinations in the second step are linearly independent of the linear combinations revealed in the first step. Essentially, the revealed linear combinations describe the positions of some auxiliary solutions. For non-zero-preserving perturbations revealing this information is not critical as no solution has in any objective function exactly the same value as  $x$ . For zero-preserving solutions it can, however, happen that the auxiliary solutions take exactly the same value as  $x$  in one of the objective functions. Then there is not enough randomness left in the second step anymore to bound the probability that  $x$  lies in this objective in the  $\varepsilon$ -interval described by the  $\varepsilon$ -box  $B$ .

In the remainder of this section we will present some more details on our analysis. We first present a simplified argument to bound the smoothed number of Pareto-optimal solutions. Afterwards we will briefly discuss which changes to this argument are necessary to bound higher moments and to analyze zero-preserving perturbations.

**Smoothed Number of Pareto-optimal Solutions** As an important building block in the proof of Theorem 3 we use an insight from [13] how it can be tested whether a given  $\varepsilon$ -box contains a Pareto-optimal solution. Let us fix an  $\varepsilon$ -box  $B = (b_1, b_1 + \varepsilon] \times \dots \times (b_d, b_d + \varepsilon]$ . The following algorithm takes as parameters the matrix  $V$  and the  $\varepsilon$ -box  $B$  and it returns a solution  $x^{(0)}$ .

**Witness**( $V, B$ )

- 1: Set  $\mathcal{R}_{d+1} = \mathcal{S}$ .
- 2: **for**  $t = d, d-1, \dots, 0$  **do**
- 3:   Set  $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} : V^{1\dots t}z \leq b|_{1\dots t}\}$ .
- 4:   Set  $x^{(t)} = \arg \min \{V^{t+1}z : z \in \mathcal{C}_t\}$ .
- 5:   Set  $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} : V^{t+1}z < V^{t+1}x^{(t)}\}$ .
- 6: **end for**
- 7: **return**  $x^{(0)}$

The actual **Witness** function that we use in the proof of Theorem 3 is more complex because it has to deal with some technicalities. In particular, the case that some set  $\mathcal{C}_t$  is empty has to be handled. For the purpose of illustration we ignore these technicalities here and assume that  $\mathcal{C}_t$  is never empty. The crucial observation that has been made by Moitra and O'Donnell is that if there is a Pareto-optimal solution  $x \in \mathcal{S}$  that lies in  $B$ , then  $x^{(0)} = x$  (assuming that no two solutions lie in the same  $\varepsilon$ -box). Hence, the solution  $x^{(0)}$  returned by the **Witness** function is the only candidate for a Pareto-optimal solution in  $B$ . While this statement and the following reasoning are true for any  $d \in \mathbb{N}$ , we recommend the reader to think first only about the illustrative case  $d = 2$ , in which there are one adversarial and two linear objective functions.

Our goal is to execute the **Witness** function and to obtain the solution  $x^{(0)}$  without revealing the entire matrix  $V$ . We will see that it is indeed possible to divide the draw of  $V$  into two steps such that in the first step enough information is revealed to execute the **Witness** function and such that in the second step there is still enough randomness left to bound the probability that  $x^{(0)}$  lies in  $B$ . For this let  $I \subseteq [n]$  be a set of indices and assume that we know in advance which values the solutions  $x^{(0)}, \dots, x^{(d)}$  take at these indices, i.e., assume that we know  $a^{(0)} = x^{(0)}|_I, \dots, a^{(d)} = x^{(d)}|_I$  before executing the **Witness** function. Then we can reconstruct  $x^{(0)}, \dots, x^{(d)}$  without having to reveal the entire matrix  $V$ . This can be done by the following algorithm, which gets as additional parameters the set  $I$  and the matrix  $A = (a^{(0)}, \dots, a^{(d)})$ .

**Witness**( $V, I, A, B$ )

- 1: Set  $\mathcal{R}_{d+1} = \mathcal{S} \cap \bigcup_{t'=0}^d \mathcal{S}_I(a^{(t')})$ .
- 2: **for**  $t = d, d-1, \dots, 0$  **do**
- 3:   Set  $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} \cap \mathcal{S}_I(a^{(t)}) : V^{1\dots t}z \leq b|_{1\dots t}\}$ .
- 4:   Set  $x^{(t)} = \arg \min \{V^{t+1}z : z \in \mathcal{C}_t\}$ .
- 5:   Set  $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} \cap \bigcup_{t'=0}^{t-1} \mathcal{S}_I(a^{(t')}) : V^{t+1}z < V^{t+1}x^{(t)}\}$ .
- 6: **end for**
- 7: **return**  $(x^{(0)}, \dots, x^{(d)})$

The additional restriction of the set  $\mathcal{R}_{d+1}$  does not change the outcome of the **Witness** function as all solutions  $x^{(0)}, \dots, x^{(d)}$  generated by the first **Witness** function are contained in the set  $\mathcal{R}_{d+1}$  defined in line 1 of the second **Witness** function. Similarly one can argue that the additional restrictions in lines 3 and 5 do not change the outcome of the algorithm because all solutions  $x^{(t)}$  generated by the first **Witness** function satisfy the restrictions that are made in the second **Witness** function. Hence, if  $a^{(0)} = x^{(0)}|_I, \dots, a^{(d)} = x^{(d)}|_I$ , then both **Witness** functions generate the same  $x^{(0)}$ .

We will now discuss how much information about  $V$  needs to be revealed in order to execute the second **Witness** function, assuming that the additional parameters  $I$  and  $A$  are given. We assume that the coefficients  $V_i^t$  are revealed for every  $t \in [d]$  and  $i \notin I$ . For the remaining coefficients only certain linear combinations need to be known in order to be able to execute the **Witness** function. By carefully looking at the **Witness** function, one can deduce that for  $t \in [d]$  only the following linear combinations need to be known:

$$V^t|_I \cdot x^{(t)}|_I, \dots, V^t|_I \cdot x^{(d)}|_I, \\ V^t|_I \cdot (x^{(t-1)} - x^{(0)})|_I, \dots, V^t|_I \cdot (x^{(t-1)} - x^{(t-2)})|_I.$$

These terms can be viewed as linear combinations of the random variables  $V_i^t$ ,  $t \in [d]$ ,  $i \in I$ , with coefficients from  $\{-1, 0, +1\}$ . In addition to the already fixed random variables  $V_i^t$ ,  $t \in [d]$ ,  $i \notin I$  the following  $d$  linear combinations determine the position  $V^{1\dots d}x$  of  $x = x^{(0)}$ :

$$V_I^1 \cdot x^{(0)}|_I, \dots, V_I^d \cdot x^{(0)}|_I.$$

An important observation on which our analysis is based is that if the vectors  $x^{(0)}|_I, \dots, x^{(d)}|_I$  are linearly independent, then also all of the above mentioned linear combinations are linearly independent. In particular, the  $d$  linear combinations that determine the position of  $x$  cannot be expressed by the other linear combinations. Usually, however, it is not possible to find a subset  $I \subseteq [n]$  of indices such that the vectors  $x^{(0)}|_I, \dots, x^{(d)}|_I$  are linearly independent. By certain technical modifications of the **Witness** function we will ensure that there always exists such a set  $I$  with  $|I| \leq d+1$ . Since we do not know the set  $I$  and the matrix  $A$  in advance, we apply a union bound over all valid choices for these parameters, which yields to a factor of  $O(n^d)$  in the bound for the probability that there exists a Pareto-optimal solution in  $B$ .

Röglin and Teng [17] observed that the linear independence of the linear combinations implies that even if the linear combinations needed to execute the **Witness** function are revealed in the

first step, there is still enough randomness in the second step to prove an upper bound on the probability that  $V^{1\dots d}x$  lies in a fixed  $\varepsilon$ -box  $B$  that is proportional to  $\varepsilon^d$ . The bound proven in [17] is, however, not strong enough to improve Moitra and O'Donnell's result [13] because it is in the order of  $\Theta(\varepsilon^d \phi^{d^2})$ , which would yield a bound of  $O(n^{2d} \phi^{d^2})$  for the smoothed number of Pareto-optimal solutions. We show that for quasiconcave density functions the bound in [17] can be improved significantly to  $\Theta(\varepsilon^d \phi^d)$ , which yields the improved bound of  $O(n^{2d} \phi^d)$  in Theorem 3.

**Higher Moments** The analysis of higher moments is based on running the **Witness** function multiple times. As described above, we bound for a fixed  $c$ -tuple  $(B_1, \dots, B_c)$  of  $\varepsilon$ -boxes the probability that all of them contain a Pareto-optimal solution. For this, we run the **Witness** function  $c$  times. This way, we get for every  $j \in [c]$  a sequence  $x^{(j,0)}, \dots, x^{(j,d)}$  of solutions such that  $x^{(j,0)}$  is the unique candidate for a Pareto-optimal solution in  $B_j$ .

As above, we would like to execute the  $c$  calls of the **Witness** function without having to reveal the entire matrix  $V$ . Again if we know for a subset  $I \subseteq [n]$  the values that the solutions  $x^{(j,t)}$ ,  $j \in [c]$ ,  $t \in [d]$ , take at these positions, then we do not need to reveal the coefficients  $V_i^t$  with  $i \in I$  to be able to execute the calls of the **Witness** function. As in the case of the first moment, it suffices to reveal some linear combinations of these coefficients.

In order to guarantee that these linear combinations are linearly independent of the linear combinations that determine the positions of the solutions  $x^{(j,0)}$ ,  $j \in [c]$ , we need to coordinate the calls of the **Witness** function. Otherwise it might happen, for example, that the linear combinations revealed for executing the first call of the witness function determine already the position of  $x^{(2,0)}$ , the candidate for a Pareto-optimal solution in  $B_2$ . Assume that the first call of the **Witness** function returns a sequence  $x^{(1,0)}, \dots, x^{(1,d)}$  of solutions and that  $I_1 \subseteq [n]$  is a set of indices that satisfies the desired property that  $x^{(1,0)}|_{I_1}, \dots, x^{(1,d)}|_{I_1}$  are linearly independent. In order to achieve that all solutions generated in the following calls of the **Witness** function are linearly independent of these linear combinations, we do not start a second independent call of the **Witness** function, but we restrict the set of feasible solutions first. Instead of choosing  $x^{(2,0)}, \dots, x^{(2,d)}$  among all solutions from  $\mathcal{S}$ , we restrict the set of feasible solutions for the second call of the **Witness** function to  $\mathcal{S}' = S_{I_1}(x)$ . Essentially, all solutions generated in calls of the **Witness** function have to coincide with  $x$  in all positions that have been selected in one of the previous calls.

This and some additional tricks allow us to ensure that in the end there is a set  $I \subseteq [n]$  with  $|I| \leq (d+1)c$  such that all vectors  $x^{(j,t)}|_I$ ,  $j \in [c]$ ,  $t \in [d]$  are linearly independent. Then we can again use the bound proven in [17] to bound the probability that  $V^{1\dots d}x^{(j,0)} \in B_j$  simultaneously for every  $j \in [c]$  from above by a term proportional to  $\varepsilon^{cd} \phi^{cd^2}$ . With our improved bound for quasiconcave density functions, we obtain a bound proportional to  $\varepsilon^{cd} \phi^{cd}$ . Together with a union bound over all valid choices for  $I$  and the values  $x^{(j,t)}|_I$ ,  $j \in [c]$ ,  $t \in [d]$ , we obtain a bound of  $\kappa n^{cd} \varepsilon^{cd} \phi^{cd}$  on the probability that all candidates  $x^{(j,0)}$  lie in their corresponding  $\varepsilon$ -boxes for a sufficiently large constant  $\kappa$ . Together with the bound of  $O(n^{cd}/\varepsilon^{cd})$  for the number of  $c$ -tuples  $(B_1, \dots, B_c)$  this implies Theorem 4.

**Zero-preserving Perturbations** If we use the same **Witness** function as above, then it can happen that there is a Pareto-optimal solution  $x$  in the  $\varepsilon$ -box  $B$  that does not coincide with the solution  $x^{(0)}$  returned by the **Witness** function. This problem occurs, for example, if  $V^d \cdot x^{(d-1)} = V^d \cdot x^{(0)}$ , which we cannot exclude if we allow zero-preserving perturbations. We recommend to visualize this case for  $d = 2$ . On the other hand if we knew in advance that  $V^d \cdot x^{(d-1)} = V^d \cdot x^{(0)}$ , then we could bound the probability of  $V^d x^{(0)} \in (b_d, d_d + \varepsilon]$  already after the solution  $x^{(d-1)}$  has been generated. Hence, if we were only interested in bounding this probability, we could terminate the **Witness** function already after  $x^{(d-1)}$  has been generated. Instead of terminating the **Witness** function at this point entirely, we keep in mind that  $V^d x^{(0)}$  has already been determined and we restart the **Witness** function with the remaining objective functions only.

Let us make this a bit more precise. As long as the solutions  $x^{(t)}$  generated by the **Witness** function differ in all objective functions from  $x$ , we execute the **Witness** function without any modification. Only if a solution  $x^{(t)}$  is generated that agrees with  $x$  in some objective functions, we deviate from the original **Witness** function. Let  $K \subseteq [d]$  denote the objective functions in which  $x^{(t)}$  coincides with  $x$ . Then we can bound at this point the probability that  $V^t \cdot x \in (b_t, b_t + \varepsilon]$  simultaneously for all  $t \in K$ . In order to also deal with the other objectives  $t \notin K$ , we restart the **Witness** function. In this restart, we ignore all objective functions in  $K$  and we execute the **Witness** function as if only objectives  $t \notin K$  were present. Additionally we restrict in the restart the set of feasible solutions to those that coincide in the objectives  $t \in K$  with  $x$ , i.e., to  $\{y \in \mathcal{S} \mid \forall t \in K : V^t \cdot y = V^t \cdot x\}$ . With similar techniques as in the analysis of higher moments we ensure that different restarts lead to linearly independent linear combinations.

This exploits that every Pareto-optimal solution  $x$  is also Pareto-optimal with respect to only the objective functions  $V^t$  with  $t \notin K$  if the set  $\mathcal{S}$  is restricted to solutions that agree with  $x$  in all objective functions  $V^t$  with  $t \in K$ . This property guarantees that whenever the **Witness** function is restarted,  $x$  is still a Pareto-optimal solution with respect to the restricted solution set and the remaining objective functions.

It can happen that we have to restart the **Witness** function  $d$  times before a unique candidate for a Pareto-optimal solution in  $B$  is identified. As in each of these restarts at most  $d$  solutions are generated, the total number of solutions that is generated can increase from  $d + 1$ , as in the case of non-zero-preserving perturbations, to roughly  $d^2$ . The set  $I \subseteq [n]$  of indices restricted to which these solutions are linearly independent has a cardinality of at most  $d^3$ . The reason for this increase is that we have to choose more indices to obtain linear independence due to the fixed zeros. Taking a union bound over all valid choices of  $I$ , of the values that the generated solutions take at these positions, and the possibilities when and due to which objectives the restarts occur, yields Theorem 1. This theorem relies again on the result about linearly independent linear combinations of independent random variables from [17] and its improved version for quasiconcave densities that we show in this paper.

## 4 Non-zero-preserving Perturbations

### 4.1 Smoothed Number of Pareto-optimal Solutions

To prove Theorem 3 we assume without loss of generality that  $n \geq d + 1$  and consider the following function called the **Witness** function. It is very similar to the one suggested by Moitra and O'Donnell, but with an additional parameter  $I$ . This parameter is a tuple of forbidden indices: it restricts the set of indices we are allowed to choose from. For the analysis of the smoothed number of Pareto-optimal solutions we will set  $I = ()$ . The parameter becomes important in the next section when we analyze higher moments.

**Witness**( $V, x, I$ )

- 1: Set  $\mathcal{R}_{d+1} = \mathcal{S} \cap \mathcal{S}_I(x)$ .
- 2: **for**  $t = d, d - 1, \dots, 0$  **do**
- 3:   Set  $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} : V^{1\dots t} z < V^{1\dots t} x\}$ .
- 4:   **if**  $\mathcal{C}_t \neq \emptyset$  **then**
- 5:     Set  $x^{(t)} = \arg \min \{V^{t+1} z : z \in \mathcal{C}_t\}$ .
- 6:     **if**  $t = 0$  **then return**  $x^{(t)}$
- 7:     Determine the first index  $i_t \in [n]$  such that  $x_{i_t}^{(t)} \neq x_{i_t}$ .
- 8:      $I \leftarrow I \cup (i_t)$
- 9:     Set  $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} \cap \mathcal{S}_I(x) : V^{t+1} z < V^{t+1} x^{(t)}\}$ .
- 10: **else**
- 11:   Set  $i_t = \min([n] \setminus I)$ .

```

12:    $I \leftarrow I \cup \{i_t\}$ .
13:   Set  $x_i^{(t)} = \begin{cases} \bar{x}_i & : i = i_t, \\ x_i & : \text{otherwise.} \end{cases}$ 
14:   Set  $\mathcal{R}_t = \mathcal{R}_{t+1} \cap \mathcal{S}_I(x)$ .
15: end if
16: end for
17: return  $(\perp, \dots, \perp)$ .

```

Note, that  $\mathcal{C}_0 = \mathcal{R}_1$  since  $V^{1\dots t}z < V^{1\dots t}x$  is no restriction if  $t = 0$ . For  $t \geq 1$  the index  $i_t$  in line 7 exists because  $V^1x^{(t)} < V^1x$ , which implies  $x^{(t)} \neq x$ . In the remainder of this section we only consider the case that  $x$  is Pareto-optimal, that  $I$  contains pairwise distinct indices and that the number  $|I|$  of indices contained in  $I$  is at most  $n - (d + 1)$ . This ensures that the indices  $i_0, \dots, i_d$  exist. Unless stated otherwise, we assume that the OK-event occurs. That means that  $|V^k \cdot (y - z)| \geq \varepsilon$  for every  $k \in [d]$  and for any two distinct vectors  $y \neq z \in \mathcal{S}$  and that for any  $k \in [d]$  there is an index  $i \in [n]$  such that  $|V_i^k| < 1$ . The first property ensures, amongst others, that there is a unique argmin in line 5. The latter property, which holds with probability 1, ensures that  $B_V(x) \in \mathbb{B}_\varepsilon$  for any vector  $x \in \{-1, 0, 1\}^n$ .

**Lemma 6.** *The call  $\text{Witness}(V, x, I)$  returns the vector  $x^{(0)} = x$ .*

*Proof.* Like Moitra and O'Donnell, we prove two claims to show Lemma 6.

**Claim 1.** *There is no  $t \in [d + 1]$  such that there exists a solution  $z \in \mathcal{R}_t$  that dominates  $x$  with respect to the objectives  $V^1, \dots, V^t$ .*

**Claim 2.** *For any  $t \in [d + 1]$  the solution  $x$  is an element of  $\mathcal{R}_t$ .*

If both claims hold, then the correctness of Lemma 6 follows immediately: Claim 2 yields  $x \in \mathcal{C}_0 = \mathcal{R}_1$  and, hence, we return  $x^{(0)} = \argmin\{V^1z : z \in \mathcal{C}_1\}$  defined in line 5. Due to Claim 1,  $x^{(0)} = x$ .

Let us first consider Claim 1. For  $t = d + 1$  the statement follows from the Pareto-optimality of  $x$ . Thus, let  $t \leq d$  and assume for contradiction that there is a vector  $z \in \mathcal{R}_t \subseteq \mathcal{R}_{t+1}$  that dominates  $x$  with respect to  $V^{1\dots t}$ . Then,  $z \neq x$  and consequently  $V^{1\dots t}z < V^{1\dots t}x$  due to the occurrence of the OK-event. Hence, we obtain  $z \in \mathcal{C}_t$  and  $V^{t+1}x^{(t)} \leq V^{t+1}z$  due to the choice of  $x^{(t)}$  in line 5. This contradicts the inequality  $V^{t+1}z < V^{t+1}x^{(t)}$  which holds since  $z \in \mathcal{R}_t$  and due to the construction of  $\mathcal{R}_t$  in line 9.

To prove Claim 2 we use downward induction over  $t$ . For  $t = d + 1$  the claim trivially holds as  $x \in \mathcal{S}$  and as  $x \in \mathcal{S}_I(x)$ . Now, let  $t \leq d$ . If  $\mathcal{C}_t = \emptyset$ , then  $x \in \mathcal{R}_t$  since  $x \in \mathcal{R}_{t+1}$  by the induction hypothesis and since  $x \in \mathcal{S}_I(x)$ . If  $\mathcal{C}_t \neq \emptyset$ , then assume for contradiction that  $x \notin \mathcal{R}_t$ . Since  $x \in \mathcal{R}_{t+1} \cap \mathcal{S}_I(x)$  due to the induction hypothesis, this yields  $V^{t+1}x^{(t)} \leq V^{t+1}x$  which contradicts Claim 1 as it implies together with  $x^{(t)} \in \mathcal{R}_{t+1}$  that  $x^{(t)}$  dominates  $x$  with respect to the objectives  $V^1, \dots, V^{t+1}$ .  $\square$

We will see that for the execution of the call  $\text{Witness}(V, x, I)$  we only need some information about  $x$ . This is one crucial property which will help us to bound the expected number of Pareto-optimal solutions.

**Definition 7.** *Let  $x^{(0)}, \dots, x^{(d)}$  be the vectors constructed during the call  $\text{Witness}(V, x, I)$ , let  $\hat{I}$  be the tuple  $I$  at the moment when the  $\text{Witness}$  function terminates and let  $i^* = \min([n] \setminus \hat{I})$ . We call the pair  $(I^*, A^*)$  where  $I^* = \hat{I} \cup \{i^*\}$  and  $A^* = [x^{(d)}, \dots, x^{(0)}]$  the  $(V, I)$ -certificate of  $x$ . The pair  $(I^*, A^*|_{I^*})$  is called the restricted  $(V, I)$ -certificate of  $x$ . We call a pair  $(I', A')$  a (restricted)  $I$ -certificate, if there exist a realization  $V$  such that  $\text{OK}(V)$  is true and a Pareto-optimal solution  $x \in \mathcal{S}$  such that  $(I', A')$  is the (restricted)  $(V, I)$ -certificate of  $x$ . By  $\mathcal{C}(I)$  we denote the set of all restricted  $I$ -certificates.*

For the analysis of the first moment we only need restricted  $I$ -certificates. Our analysis of higher moments requires more knowledge about the vectors  $x^{(t)}$  than just the values  $x_i$  for  $i \in I^*$ . The additional indices are, however, depending on further calls of the **Witness** function which we do not know a priori. This is why we have to define two types of certificates. For the sake of reusability we formulate some statements more general than necessary for this section.

**Lemma 8.** *Let  $V$  be an arbitrary realization where  $\text{OK}(V)$  is true, let  $x$  be a Pareto-optimal solution with respect to  $V$ , and let  $(I^*, A)$  be the restricted  $(V, I)$ -certificate of  $x$ . Then,  $I^* = (j_1, \dots, j_{|I|+d+1})$  consists of pairwise distinct indices and*

$$A = \begin{bmatrix} x_{j_1} & \dots & x_{j_{|I|}} & \overline{x_{j_{|I|+1}}} & * & \dots & * \\ \vdots & & \vdots & x_{j_{|I|+1}} & \ddots & \ddots & \vdots \\ \vdots & & \vdots & \vdots & \ddots & \overline{x_{j_{|I|+d}}} & * \\ x_{j_1} & \dots & x_{j_{|I|}} & x_{j_{|I|+1}} & \dots & x_{j_{|I|+d}} & x_{j_{|I|+d+1}} \end{bmatrix}^T \in \{0, 1\}^{(|I|+d+1) \times (d+1)},$$

where each ‘ $*$ ’ can either be 0 or 1 independently of the other ‘ $*$ ’-entries.

*Proof.* Lemma 6 implies that the last column of  $A$  equals  $x|_{I^*}$ . Hence, we just have to consider the first  $d$  columns of  $A$ . Note, that  $I = (j_1, \dots, j_{|I|})$  and  $j_{|I|+1}, \dots, j_{|I|+d+1} = i_d, \dots, i_1, i^*$ . Let  $I_t = I \cup (i_d, \dots, i_t)$ . The construction of the sets  $\mathcal{R}_t$  yields  $\mathcal{R}_t \subseteq \mathcal{S}_{I_t}(x)$  (see lines 1, 9, and 14). Index  $i_t$  is always chosen such that  $i_t \notin I_{t+1}$ : If it is constructed in line 7, then  $x_{i_t}^{(t)} \neq x_{i_t}$ . Since in that case we have  $x^{(t)} \in \mathcal{R}_{t+1} \subseteq \mathcal{S}_{I_{t+1}}(x)$  and, hence, index  $i_t$  cannot be an element of  $I_{t+1}$ . In line 11, index  $i_t$  is explicitly constructed such that  $i_t \notin I_{t+1}$ . The same argument holds for the index  $i^*$ . Hence, the indices of  $I^*$  are pairwise distinct.

Now, consider the column of  $A$  corresponding to vector  $x^{(t)}$  for  $t \in [d]$ . If  $\mathcal{C}_t = \emptyset$ , then the form of the column follows directly from the construction of  $x^{(t)}$  in line 13 and from the fact that the indices of  $I^*$  are pairwise distinct. If  $\mathcal{C}_t \neq \emptyset$ , then  $x^{(t)} \in \mathcal{C}_t \subseteq \mathcal{R}_{t+1} \subseteq \mathcal{S}_{I_{t+1}}(x)$ , i.e.,  $x^{(t)}$  agrees with  $x$  in all indices  $i \in I_{t+1}$ . By the choice of  $i_t$  in line 7 we get  $x_{i_t}^{(t)} = \overline{x_{i_t}}$ . This concludes the proof.  $\square$

Let  $(I^*, A^*)$  be the  $(V, I)$ -certificate of  $x$  and let  $J \supseteq I^*$  be a tuple of pairwise distinct indices. We consider the following variant of the **Witness** function that uses information about  $x$  given by the index tuple  $J$ , the bit matrix  $A = A^*|_J$  with columns  $a^{(d)}, \dots, a^{(0)}$ , a shift vector  $u$  and the  $\varepsilon$ -box  $B = B_V(x - u)$  instead of vector  $x$  itself. The meaning of the shift vector will become clear when we analyze the probability of certain events. We will see that not all information about  $V$  needs to be revealed to execute the new **Witness** function, i.e., we have some randomness left which we can use later. With the choice of the shift vector we can control which information needs to be revealed for executing the **Witness** function.

**Witness** $(V, J, A, B, u)$

- 1: Let  $b$  be the corner of  $B$ .
- 2: Set  $\mathcal{R}_{d+1} = \mathcal{S} \cap \bigcup_{t'=0}^d \mathcal{S}_J(a^{(t')})$ .
- 3: **for**  $t = d, d-1, \dots, 0$  **do**
- 4:   Set  $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} \cap \mathcal{S}_J(a^{(t)}) : V^{1\dots t} \cdot (z - u) \leq b|_{1\dots t}\}$ .
- 5:   **if**  $\mathcal{C}_t \neq \emptyset$  **then**
- 6:     Set  $x^{(t)} = \arg \min \{V^{t+1}z : z \in \mathcal{C}_t\}$ .
- 7:     **if**  $t = 0$  **then return**  $x^{(t)}$
- 8:     Set  $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} \cap \bigcup_{t'=0}^{t-1} \mathcal{S}_J(a^{(t')}) : V^{t+1}z < V^{t+1}x^{(t)}\}$ .
- 9:   **else**
- 10:     Set  $x^{(t)} = (\perp, \dots, \perp)$ .
- 11:     Set  $\mathcal{R}_t = \mathcal{R}_{t+1} \cap \bigcup_{t'=0}^{t-1} \mathcal{S}_J(a^{(t')})$ .

12: **end if**  
 13: **end for**  
 14: **return**  $(\perp, \dots, \perp)$

**Lemma 9.** *Let  $(I^*, A^*)$  be the  $(V, I)$ -certificate of  $x$ , let  $J \supseteq I^*$  be an arbitrary tuple of pairwise distinct indices, let  $A = A^*|_J$ , let  $u \in \{0, 1\}^n$  be an arbitrary vector, and let  $B = B_V(x - u)$ . Then, the call  $\text{Witness}(V, J, A, B, u)$  returns vector  $x$ .*

Before we give a formal proof of Lemma 9 we try to give some intuition for it. Instead of considering the whole set  $\mathcal{S}$  of solutions we restrict it to vectors that look like the vectors we want to reconstruct in the next rounds, i.e., we intersect the current set with the set  $\bigcup_{t'=0}^{t-1} \mathcal{S}_J(a^{(t')})$  in round  $t$ . That way we only deal with subsets of the original sets, but we do not lose the vectors we want to reconstruct since  $J \supseteq I^*$ .

*Proof.* Let  $\mathcal{R}'_t, \mathcal{C}'_t$ , and  $x^{(t)}$  denote the sets and vectors constructed during the execution of the call  $\text{Witness}(V, J, A, B, u)$  and let  $\mathcal{R}_t, \mathcal{C}_t$ , and  $x^{(t)}$  denote the sets and vectors constructed during the execution of call  $\text{Witness}(V, x, I)$ . By induction we prove the following statements:

- $\mathcal{R}'_t \subseteq \mathcal{R}_t$  for any  $t \in [d+1]$ .
- $x^{(t)} = x^{(t)}$  for any  $t \in [d]_0$  for which  $\mathcal{C}_t \neq \emptyset$ .
- $x^{(t')} \in \mathcal{R}'_t$  for any  $t \in [d+1]$  and any  $t' \in [t-1]$  for which  $\mathcal{C}_{t'} \neq \emptyset$ .

With those claims Lemma 9 follows immediately: Since  $x^{(0)} = x$  and  $\mathcal{C}_0 \neq \emptyset$  due to Lemma 6 we obtain  $x^{(0)} \in \mathcal{R}'_1 = \mathcal{C}'_0$  and, hence, we return  $x^{(0)} = x^{(0)} = x$ .

Let us first focus on the shift vector  $u$  and compare line 3 of the first **Witness** function with line 4 of the second **Witness** function. The main difference is that in the first version we have the restriction  $V^{1\dots t}z < V^{1\dots t}x$ , whereas in the second version we seek for solutions  $z$  such that  $V^{1\dots t} \cdot (z - u) \leq b|_{1\dots t}$ . As  $b$  is the corner of the  $\varepsilon$ -box  $B = B_V(x - u)$ , those restrictions are equivalent for solutions  $z \in \mathcal{S}$  since

$$V^{1\dots t} \cdot (z - u) \leq b|_{1\dots t} \iff V^{1\dots t} \cdot (z - u) < V^{1\dots t} \cdot (x - u) \iff V^{1\dots t}z < V^{1\dots t}x.$$

The first inequality is due to the occurrence of the OK-event. Now, we prove the statements by downward induction over  $t$ . Let  $t = d+1$ . Lemma 8 yields  $a^{(t')}_I = x|_I$  for any  $t' \in [d]_0$ , i.e.,  $\bigcup_{t'=0}^d \mathcal{S}_J(a^{(t')}) \subseteq \mathcal{S}_I(x)$ . Consequently,  $\mathcal{R}'_{d+1} \subseteq \mathcal{R}_{d+1}$ . Consider an arbitrary index  $t' \in [d]$  for which  $\mathcal{C}_{t'} \neq \emptyset$ . Then,  $x^{(t')} \in \mathcal{R}_{t'} \subseteq \mathcal{S}$  and  $x^{(t')} \in \mathcal{S}_J(a^{(t')})$ . Hence,  $x^{(t')} \in \mathcal{R}'_{d+1}$ .

Let  $t \leq d$ . By the observation above we have  $\mathcal{C}'_t = \{z \in \mathcal{R}'_{t+1} \cap \mathcal{S}_J(a^{(t)}) : V^{1\dots t}z < V^{1\dots t}x\}$  and  $\mathcal{C}_t = \{z \in \mathcal{R}_{t+1} : V^{1\dots t}z < V^{1\dots t}x\}$ . Since  $\mathcal{R}'_{t+1} \subseteq \mathcal{R}_{t+1}$ , we obtain  $\mathcal{C}'_t \subseteq \mathcal{C}_t$ . We first consider the case  $\mathcal{C}_t = \emptyset$ , which implies  $\mathcal{C}'_t = \emptyset$ . Then,  $\mathcal{R}'_t = \mathcal{R}'_{t+1} \cap \bigcup_{t'=0}^{t-1} \mathcal{S}_J(a^{(t')})$  and  $\mathcal{R}_t = \mathcal{R}_{t+1} \cap \mathcal{S}_I(x)$  for the current index tuple  $I$ . According to Lemma 8, all vectors  $x^{(0)}, \dots, x^{(t-1)}$  agree with  $x$  on the indices  $i \in I$ . Thus,  $\bigcup_{t'=0}^{t-1} \mathcal{S}_J(a^{(t')}) \subseteq \mathcal{S}_I(x)$ , and as  $\mathcal{R}'_{t+1} \subseteq \mathcal{R}_{t+1}$  we obtain  $\mathcal{R}'_t \subseteq \mathcal{R}_t$ . Let  $t' \in [t-1]$  be an index for which  $\mathcal{C}_{t'} \neq \emptyset$ . Then,  $x^{(t')} \in \mathcal{R}'_{t+1}$  by the induction hypothesis,  $x^{(t')} \in \mathcal{S}_J(a^{(t')})$ , and consequently  $x^{(t')} \in \mathcal{R}'_t$ .

Finally, let us consider the case  $\mathcal{C}_t \neq \emptyset$ . The induction hypothesis yields  $x^{(t)} \in \mathcal{R}'_{t+1}$ . Since  $x^{(t)} \in \mathcal{S}_J(a^{(t)})$  and  $V^{1\dots t} \cdot x^{(t)} < V^{1\dots t}x$ , also  $x^{(t)} \in \mathcal{C}'_t$  and thus  $\mathcal{C}'_t \neq \emptyset$ . Hence,  $x^{(t)} = x^{(t)}$  as  $\mathcal{C}'_t \subseteq \mathcal{C}_t$ . The remaining claims have to be validated only if  $t \geq 1$ . Then,  $\mathcal{R}'_t = \{z \in \mathcal{R}'_{t+1} \cap \bigcup_{t'=0}^{t-1} \mathcal{S}_J(a^{(t')}) : V^{t+1}z < V^{t+1}x^{(t)}\}$  and  $\mathcal{R}_t = \{z \in \mathcal{R}_{t+1} \cap \mathcal{S}_I(x) : V^{t+1}z < V^{t+1}x^{(t)}\}$  for the current index tuple  $I$ . With the same argument used for the case  $\mathcal{C}_t = \emptyset$  we obtain  $\mathcal{R}'_{t+1} \cap \bigcup_{t'=0}^{t-1} \mathcal{S}_J(a^{(t')}) \subseteq \mathcal{R}_{t+1} \cap \mathcal{S}_I(x)$  and thus  $\mathcal{R}'_t \subseteq \mathcal{R}_t$ . Consider an arbitrary index  $t' \in [t-1]$  for which  $\mathcal{C}_{t'} \neq \emptyset$ . Then,  $x^{(t')} \in \mathcal{R}_{t'} \subseteq \mathcal{R}_t$ . In particular,  $V^{t+1}x^{(t')} < V^{t+1}x^{(t)}$ . Furthermore,  $x^{(t')} \in \mathcal{R}'_{t+1}$  due to the induction hypothesis, and  $x^{(t')} \in \mathcal{S}_J(a^{(t')})$ . Consequently,  $x^{(t')} \in \mathcal{R}'_t$ .  $\square$

As mentioned earlier, with the shift vector  $u$  we control which information of  $V$  has to be revealed to execute the call  $\text{Witness}(V, J, A, B, u)$ . While Lemma 9 holds for any vector  $u$ , we have to choose  $u$  carefully for our probabilistic analysis to work. We will see that the choice  $u^* = u^*(J, A)$ ,

$$u_i^* = \begin{cases} \overline{x_i} & : i = i^*, \\ x_i & : i \in J \setminus (i^*), \\ 0 & : \text{otherwise,} \end{cases} \quad (1)$$

is appropriate. Note, that  $i^*$  is the index that has been added to  $\hat{I}$  in the definition of the  $(V, I)$ -certificate to obtain  $I^*$ . Moreover, the values  $x_i$  are encoded in the last column of  $A$  for any index  $i \in J$  according to Lemma 8. Hence, if  $(I^*, A^*)$  is the  $(V, I)$ -certificate of  $x$ , then vector  $u^*$  can be defined when a tuple  $J \supseteq I^*$  and the matrix  $A = A^*|_J$  are known; we do not have to know the vector  $x$  itself.

For bounding the number of Pareto-optimal solutions consider the functions  $\chi_{I^*, A, B}(V)$  parameterized by an arbitrary restricted  $I$ -certificate  $(I^*, A)$ , and an arbitrary  $\varepsilon$ -box  $B \in \mathbb{B}_\varepsilon$ , defined as follows:  $\chi_{I^*, A, B}(V) = 1$  if the call  $\text{Witness}(V, I^*, A, B, u^*(J, A))$  returns a solution  $x' \in \mathcal{S}$  such that  $B_V(x' - u^*(I^*, A)) = B$ , and  $\chi_{I^*, A, B}(V) = 0$  otherwise.

**Corollary 10.** *Assume that  $\text{OK}(V)$  is true. Then, the number  $\text{PO}(V)$  of Pareto-optimal solutions is at most  $\sum_{(I^*, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I^*, A, B}(V)$ .*

*Proof.* Let  $x$  be a Pareto-optimal solution, let  $(I^*, A)$  be the restricted  $(V, I)$ -certificate of  $x$ , and let  $B = B_V(x - u^*(I^*, A)) \in \mathbb{B}_\varepsilon$ . Due to Lemma 9,  $\text{Witness}(V, I^*, A, B, u^*(I^*, A))$  returns vector  $x$ . Hence,  $\chi_{I^*, A, B}(V) = 1$ . It remains to show that the function  $x \mapsto (I^*, A, B')$  is injective. Let  $x_1$  and  $x_2$  be distinct Pareto-optimal solutions and let  $(I_1^*, A_1)$  and  $(I_2^*, A_2)$  be the restricted  $(V, I)$ -certificates of  $x_1$  and  $x_2$ , respectively. If  $(I_1^*, A_1) \neq (I_2^*, A_2)$ , then  $x_1$  and  $x_2$  are mapped to distinct triplets. Otherwise,  $u^*(I_1^*, A_1) = u^*(I_2^*, A_2)$  and, hence,  $B_V(x_1 - u^*(I_1^*, A_1)) \neq B_V(x_2 - u^*(I_2^*, A_2))$  because  $\text{OK}(V)$  is true. Consequently, also in this case  $x_1$  and  $x_2$  are mapped to distinct triplets.  $\square$

Corollary 10 immediately implies a bound on the expected number of Pareto-optimal solutions.

**Corollary 11.** *The expected number of Pareto-optimal solutions is bounded by*

$$\mathbf{E}_V[\text{PO}(V)] \leq \sum_{(I^*, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \mathbf{Pr}_V[E_{I^*, A, B}] + 2^n \cdot \mathbf{Pr}_V[\overline{\text{OK}(V)}]$$

where  $E_{I^*, A, B}$  denotes the event that the call  $\text{Witness}(V, I^*, A, B, u^*(I^*, A))$  returns a vector  $x'$  such that  $B_V(x' - u^*(I^*, A)) = B$ .

*Proof.* Using Corollary 10 we obtain

$$\begin{aligned} \mathbf{E}_V[\text{PO}(V)] &= \mathbf{E}_V[\text{PO}(V) \mid \text{OK}(V)] \cdot \mathbf{Pr}_V[\text{OK}(V)] + \mathbf{E}_V[\text{PO}(V) \mid \overline{\text{OK}(V)}] \cdot \mathbf{Pr}_V[\overline{\text{OK}(V)}] \\ &\leq \mathbf{E}_V \left[ \sum_{(I^*, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I^*, A, B}(V) \mid \text{OK}(V) \right] \cdot \mathbf{Pr}_V[\text{OK}(V)] + 2^n \cdot \mathbf{Pr}_V[\overline{\text{OK}(V)}] \\ &\leq \mathbf{E}_V \left[ \sum_{(I^*, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I^*, A, B}(V) \right] + 2^n \cdot \mathbf{Pr}_V[\overline{\text{OK}(V)}] \\ &= \sum_{(I^*, A) \in \mathcal{C}(I)} \sum_{B \in \mathbb{B}_\varepsilon} \mathbf{Pr}_V[E_{I^*, A, B}] + 2^n \cdot \mathbf{Pr}_V[\overline{\text{OK}(V)}]. \end{aligned} \quad \square$$



We will see that the first term of the sum in Corollary 11 can be bounded independently of  $\varepsilon$  and that the limit of the second term is 0 for  $\varepsilon \rightarrow 0$ . First of all, we analyze the size of the restricted certificate space.

**Lemma 12.** *The size of the restricted certificate space  $\mathcal{C}(I_0)$  for  $I_0 = ()$  is bounded by  $|\mathcal{C}(I_0)| \leq 2^{(d+1)^2} \cdot n^d = O(n^d)$  for fixed  $d$ .*

*Proof.* Given an index tuple  $I_0$ , exactly  $d$  indices  $i_1, \dots, i_d$  are created in  $\text{Witness}(V, x, I_0)$  if  $\text{OK}(V)$  is true and if  $x$  is Pareto-optimal with respect to  $V$ . The index  $i^*$  is determined deterministically depending on the indices  $i_1, \dots, i_d$ . Matrix  $A$  of any restricted  $I_0$ -certificate  $(I^*, A)$  is a binary  $(d+1) \times (d+1)$ -matrix. Hence, the number of possible restricted  $I_0$ -certificates is bounded by  $2^{(d+1)^2} \cdot n^d$ .  $\square$

Let us now fix an arbitrary  $I$ -certificate  $(I^*, A^*)$ , a tuple  $J \supseteq I^*$ , and an  $\varepsilon$ -box  $B \in \mathbb{B}_\varepsilon$ . We want to analyze the probability  $\Pr_V[E_{J,A,B}]$  where  $A = A^*|_J$ . By  $V_J$  and  $V_{\bar{J}}$  we denote the part of the matrix  $V^{1\dots d}$  that belongs to the indices  $i \in J$  and to the indices  $i \notin J$ , respectively. For our analysis let us further assume that the  $V_{\bar{J}}$  is fixed as well, i.e., we will only exploit the randomness of  $V_J$ .

As motivated above, the call  $\text{Witness}(V, J, A, B, u)$  can be executed without the full knowledge of  $V_J$ . To formalize this, we introduce matrices  $Q_k$  that describe the linear combinations of  $V_J^k$  we have to know:

$$Q_k = [p^{(d)}, \dots, p^{(k)}, p^{(k-2)} - p^{(k-1)}, \dots, p^{(0)} - p^{(k-1)}] \quad (2)$$

for  $p^{(t)} = p^{(t)}(J, A, u) = a^{(t)} - u|_J$  where  $a^{(t)}$  are the columns of matrix  $A$ . Note, that the matrices  $Q_k = Q_k(J, A, u)$  depend on the pair  $(J, A)$  and on the vector  $u$ .

**Lemma 13.** *Let  $u \in \{0, 1\}^n$  be an arbitrary shift vector and let  $U$  and  $W$  be two realizations of  $V$  such that  $U_{\bar{J}} = W_{\bar{J}}$  and  $U_J^k \cdot q = W_J^k \cdot q$  for any index  $k \in [d]$  and any column  $q$  of the matrix  $Q_k(J, A, u)$ . Then, the calls  $\text{Witness}(U, J, A, B, u)$  and  $\text{Witness}(W, J, A, B, u)$  return the same result.*

*Proof.* We fix an index  $k \in [d]$  and analyze which information of  $V_J^k$  is required in the call  $\text{Witness}(V, J, A, B, u)$ . For the execution of line 4 we need to know  $V_J^k \cdot (z - u)$  for solutions  $z \in \mathcal{S}_J(a^{(t)})$  in all rounds  $t \geq k$ . Since we can assume  $V_J^k$  to be known, this means that  $V_J^k \cdot (z|_J - u|_J) = V_J^k \cdot (a^{(t)} - u|_J) = V_J^k \cdot p^{(t)}$  must be revealed. For  $t \geq k$  vector  $p^{(t)}$  is a column of  $Q_k$ . The execution of line 6 does not require further information about  $V_J^k$  as all solutions in  $\mathcal{C}_t$  agree on the indices  $i \in J$ . Hence, it remains to consider line 8. Only in round  $t = k - 1$  we need information about  $V_J^k$ . In that round it suffices to know  $V_J^k \cdot (z|_J - x^{(t)}|_J)$  for any solution  $z \in \bigcup_{t'=0}^{t-1} \mathcal{S}_J(a^{(t')})$ . Hence, for  $z \in \mathcal{S}_J(a^{(t')})$ ,  $t' \in [k-2]_0$ , the linear combinations  $V_J^k \cdot (z|_J - x^{(t)}|_J) = V_J^k \cdot ((a^{(t')} - u|_J) - (a^{(k-1)} - u|_J)) = V_J^k \cdot (p^{(t')} - p^{(k-1)})$  must be revealed. Since  $t' \in [k-2]_0$ , this is a column of  $Q_k$ . As  $U$  and  $W$  agree on all necessary information, both calls return the same result.  $\square$

We will now see why  $u^* = u^*(J, A)$  defined in Equation (1) is a good shift vector.

**Lemma 14.** *Let  $Q = [\hat{p}^{(d)}, \dots, \hat{p}^{(0)}]$  where  $\hat{p}^{(t)} = p^{(t)}(J, A, u^*(J, A))|_{I^*}$ . Then,*

$$|Q| = \begin{bmatrix} 0 & \dots & 0 & 1 & * & \dots & * \\ \vdots & & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \vdots & \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 \end{bmatrix}^T \in \{0, 1\}^{(|I|+d+1) \times (d+1)},$$

where  $|Q|$  denotes the matrix  $Q'$  for which  $q'_{ij} = |q_{ij}|$ .

*Proof.* Let  $I^* = (j_1, \dots, j_{|I|+d+1})$ , i.e.,  $i^* = j_{|I|+d+1}$ . According to Lemma 8 and the construction of vector  $u^*$  in Equation (1) we obtain

$$Q = \begin{bmatrix} x_{j_1} & \dots & \dots & x_{j_1} \\ \vdots & & & \vdots \\ \overline{x_{j_{|I|}}} & \dots & \dots & x_{j_{|I|}} \\ \overline{x_{j_{|I|+1}}} & x_{j_{|I|+1}} & \dots & x_{j_{|I|+1}} \\ * & \ddots & \ddots & \vdots \\ \vdots & \ddots & \overline{x_{j_{|I|+d}}} & x_{j_{|I|+d}} \\ * & \dots & * & x_{j_{|I|+d+1}} \end{bmatrix} - \begin{bmatrix} x_{j_1} & \dots & x_{j_1} \\ \vdots & & \vdots \\ x_{j_{|I|}} & \dots & x_{j_{|I|}} \\ x_{j_{|I|+1}} & \dots & x_{j_{|I|+1}} \\ \vdots & & \vdots \\ \overline{x_{j_{|I|+d}}} & \dots & \overline{x_{j_{|I|+d}}} \\ \overline{x_{j_{|I|+d+1}}} & \dots & \overline{x_{j_{|I|+d+1}}} \end{bmatrix}$$

The claim follows since  $a - a = 0$  and  $a - \bar{a} \in \{-1, 1\}$  for  $a \in \{0, 1\}$ .  $\square$

**Lemma 15.** *For any  $k \in [d]$  the columns of the matrix  $Q_k(J, A, u^*(J, A))$  and the vector  $p^{(0)}$  are linearly independent.*

*Proof.* It suffices to show that the columns of the submatrix  $\hat{Q}_k = Q_k|_{I^*}$  and the vector  $\hat{p}^{(0)} = p^{(0)}|_{I^*}$  are linearly independent. For this, we introduce the vectors  $\hat{p}^{(t)} = p^{(t)}|_{I^*}$ . Consider the matrix  $Q = [\hat{p}^{(d)}, \dots, \hat{p}^{(0)}]$ . Due to Lemma 14 the last  $d+1$  rows of  $Q$  form a lower triangular matrix and the entries on the principal diagonal are from the set  $\{-1, 1\}$ . Consequently, the vectors  $\hat{p}^{(t)}$  are linearly independent. As these vectors are the same as the columns of matrix  $\hat{Q}_1$  plus vector  $\hat{p}^{(0)}$ , the claim holds for  $k = 1$ . Now let  $k \geq 2$ . We consider an arbitrary linear combination of the columns of  $\hat{Q}_k$  and the vector  $\hat{p}^{(0)}$  that computes to zero:

$$\begin{aligned} 0 &= \sum_{t=k}^d \lambda_t \cdot \hat{p}^{(t)} + \sum_{t=0}^{k-2} \lambda_t \cdot (\hat{p}^{(t)} - \hat{p}^{(k-1)}) + \mu \cdot \hat{p}^{(0)} \\ &= \sum_{t=k}^d \lambda_t \cdot \hat{p}^{(t)} + \sum_{t=1}^{k-2} \lambda_t \cdot \hat{p}^{(t)} - \left( \sum_{t=0}^{k-2} \lambda_t \right) \cdot \hat{p}^{(k-1)} + (\lambda_0 + \mu) \cdot \hat{p}^{(0)}. \end{aligned}$$

As the vectors  $\hat{p}^{(t)}$  are linearly independent, we first get  $\lambda_t = 0$  for  $t \in [d] \setminus \{k-1\}$ , which yields  $\lambda_0 = 0$  and, finally,  $\mu = 0$ . This concludes the proof.  $\square$

**Corollary 16.** *For an arbitrary restricted  $I$ -certificate  $(I^*, A)$  the probability of the event  $E_{I^*, A, B}$  is bounded by*

$$\Pr_V[E_{I^*, A, B}] \leq 2^d \cdot \gamma^{\gamma-d} \cdot \phi^d \cdot \varepsilon^d$$

*if all densities are quasiconcave, where  $\gamma = d \cdot (|I| + d + 1)$ .*

*Proof.* Event  $E_{I^*, A, B}$  occurs if the output is a vector  $x'$  such that  $B_V(x' - u^*(I^*, A)) = B$ . In accordance with Lemma 13 the output vector  $x'$  is determined when for any  $k \in [d]$  the linear combinations of  $V_{I^*}^k$  given by the columns of matrix  $Q_k$  are known. The equality  $B_V(x' - u^*(I^*, A)) = B$  holds if and only if

$$V^k \cdot (x' - u^*(I^*, A)) = V_{I^*}^k \cdot (x' - u^*(I^*, A))|_{\overline{I^*}} + V_{I^*}^k \cdot (x' - u^*(I^*, A))|_{I^*} \in B$$

holds for any  $k \in [d]$ . Since  $x'$  is determined and as  $(x' - u^*(I^*, A))|_{I^*} = a^{(0)} - u^*(I^*, A)|_{I^*} = p^{(0)}$  for the vector  $p^{(0)} = p^{(0)}(I^*, A, u^*(I^*, A))$ , this is equivalent to the event that

$$V_{I^*}^k \cdot p^{(0)} \in B - V_{I^*}^k \cdot (x' - u^*(I^*, A))|_{\overline{I^*}} =: C_k,$$

where  $C_k$  is an interval of length  $\varepsilon$  depending on  $x'$  and hence on the linear combinations of  $V_{I^*}$  given by the matrices  $Q_k$ . We do not explicitly mention the dependence on  $V_{\overline{I^*}}$  as we consider it

to be fixed. By  $C$  we denote the  $d$ -dimensional hypercube  $C = \prod_{k=1}^d C_k$  with side length  $\varepsilon$  defined by the intervals  $C_k$ .

For any  $k \in [d]$  let  $Q'_k$  be the matrix consisting of the columns of  $Q_k$  and the vector  $p^{(0)}$ . This matrices form the diagonal blocks of the matrix

$$Q' = \begin{bmatrix} Q'_1 & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O} \\ \mathbb{O} & \dots & \mathbb{O} & Q'_d \end{bmatrix} \in \{-1, 0, 1\}^{d \cdot |I^*| \times d \cdot (d+1)}.$$

Lemma 15 implies that matrix  $Q'$  has full rank. We permute the columns of  $Q'$  to obtain a matrix  $Q$  where the last  $d$  columns belong to the last column of one of the matrices  $Q'_k$ . Let the rows of  $Q$  be labeled by  $Q_{j_1,1}, \dots, Q_{j_m,1}, \dots, Q_{j_1,d}, \dots, Q_{j_m,d}$  assuming that  $I^* = (j_1, \dots, j_m)$ . We introduce random variables  $X_{j,k} = V_j^k$ ,  $j \in I^*$ ,  $k \in [d]$ , labeled in the same fashion as the rows of  $Q$ . Event  $E_{I^*,A,B}$  holds if and only if the  $d$  linear combinations of the variables  $X_{j,k}$  given by the last  $d$  columns of  $Q$  fall into the  $d$ -dimensional hypercube  $C$  depending on the linear combinations of the variables  $X_{j,k}$  given by the remaining columns of  $Q$ . The claim follows by applying Theorem 35 for the matrix  $Q^T$  and due to the fact that the number of rows of  $Q$  is  $d \cdot |I^*| = \gamma$ .  $\square$

*Proof of Theorem 3.* We begin the proof by showing that the OK-event is likely to happen. For any index  $t \in [d]$  and any solutions  $x \neq y \in \mathcal{S}$  the probability that  $|V^t x - V^t y| \leq \varepsilon$  is bounded by  $2\phi\varepsilon$ . To see this, choose one index  $i \in [n]$  such that  $x_i \neq y_i$  and fix all coefficients  $V_j^t$  for  $j \neq i$ . Then, the value  $V_i^t$  must fall into an interval of length  $2\varepsilon$ . A union bound over all indices  $t \in [d]$  and over all pairs  $(x, y) \in \mathcal{S} \times \mathcal{S}$  such that  $x \neq y$  yields  $\Pr_V[\overline{\text{OK}}(V)] \leq 2^{2n+1} d\phi\varepsilon$ .

With  $I_0 = ()$  and  $\gamma = d \cdot (d+1)$  we obtain

$$\begin{aligned} \mathbf{E}_V[\text{PO}(V)] &\leq \sum_{(I^*, A) \in \mathcal{C}(I_0)} \sum_{B \in \mathbb{B}_\varepsilon} \Pr_V[E_{I^*, A, B}] + 2^n \cdot \Pr_V[\overline{\text{OK}}(V)] \\ &\leq \sum_{(I^*, A) \in \mathcal{C}(I_0)} \sum_{B \in \mathbb{B}_\varepsilon} 2^d \cdot \gamma^{\gamma-d} \cdot \phi^d \cdot \varepsilon^d + 2^n \cdot 2^{2n+1} d\phi\varepsilon \\ &\leq \sum_{(I^*, A) \in \mathcal{C}(I_0)} \sum_{B \in \mathbb{B}_\varepsilon} 2^d \cdot (d+1)^{2d^2} \cdot \phi^d \cdot \varepsilon^d + 2^{3n+1} d\phi\varepsilon \\ &\leq 2^{(d+1)^2} \cdot n^d \cdot \left(\frac{2n}{\varepsilon}\right)^d \cdot 2^d \cdot (d+1)^{2d^2} \cdot \phi^d \cdot \varepsilon^d + 2^{3n+1} d\phi\varepsilon \\ &\leq 2^{(d+2)^2} \cdot (d+1)^{2d^2} \cdot n^{2d} \cdot \phi^d + 2^{3n+1} d\phi\varepsilon. \end{aligned}$$

The first inequality is due to Corollary 11. The second inequality is due to Corollary 16. The fourth inequality stems from Lemma 12. Since this bound is true for arbitrarily small  $\varepsilon > 0$ , the correctness of Theorem 3 follows.  $\square$

## 4.2 Higher Moments

The basic idea behind our analysis of higher moments is the following: If the OK-event occurs, then we can count the  $c^{\text{th}}$  power of the number  $\text{PO}(V)$  of Pareto-optimal solutions by counting all  $c$ -tuples  $(B_1, \dots, B_c)$  of  $\varepsilon$ -boxes where each  $\varepsilon$ -box  $B_i$  contains a Pareto-optimal solution  $x_i$ . We can bound this value as follows: First, call  $\text{Witness}(V, x_1, ())$  to obtain a vector  $x'_1$  and consider an index tuple  $I_1$  containing all indices created in this call and one additional index. In the second step, call  $\text{Witness}(V, x_2, I_1)$  to obtain a vector  $x'_2$  and consider the tuple  $I_2$  consisting of the indices of  $I_1$ , the indices created in this call, and one additional index. Now, call  $\text{Witness}(V, x_3, I_2)$  and

so on. If  $(x_1, \dots, x_c)$  is a tuple of Pareto-optimal solutions with  $V^{1\dots d}x_i \in B_i$  for  $i \in [c]$ , then  $(x'_1, \dots, x'_c) = (x_1, \dots, x_c)$  due to Lemma 6. As in the analysis of the first moment, we use the variant of the **Witness** function that uses certificates of the vectors  $x_\ell$  instead of the vectors itself to simulate the calls. Hence we can reuse several statements of Subsection 4.1.

Unless stated otherwise, let  $V$  be a realization such that  $\text{OK}(V)$  is true and fix arbitrary solutions  $x_1, \dots, x_c \in \mathcal{S}$  with  $V^{1\dots d}x_i \in B_i$  for  $i \in [c]$  that are Pareto-optimal with respect to  $V$ .

**Definition 17.** Let  $I_0^* = ()$  and let  $(I_\ell^*, A_\ell^*)$  be the  $(V, I_{\ell-1}^*)$ -certificate of  $x_\ell$  defined in Definition 7,  $\ell = 1, \dots, c$ . We call the pair  $(I^*, \vec{A})$ , where  $I^* = I_c^*$ ,  $\vec{A} = (A_1, \dots, A_c)$  for  $A_\ell = A_\ell^*|_{I^*}$ , the  $V$ -certificate of  $(x_1, \dots, x_c)$ . We call a pair  $(I', \vec{A}')$  a  $c$ -certificate, if there is a realization  $V$  such that  $\text{OK}(V)$  is true and if there are Pareto-optimal solutions  $x_1, \dots, x_c \in \mathcal{S}$  such that  $(I', \vec{A}')$  is the  $V$ -certificate of  $(x_1, \dots, x_c)$ . By  $\mathcal{C}_c$  we denote the set of all  $c$ -certificates.

Note, that  $I_0^* \subseteq \dots \subseteq I_c^*$  and  $|I_\ell^*| = |I_{\ell-1}^*| + d + 1$  for  $\ell \in [c]$ .

We now consider the functions  $\chi_{I^*, \vec{A}, \vec{B}}(V)$ , parameterized by an arbitrary  $c$ -certificate  $(I^*, \vec{A}) \in \mathcal{C}_c$  and a vector  $\vec{B} \in \mathbb{B}_\varepsilon^c$  of  $\varepsilon$ -boxes and defined as follows:  $\chi_{I^*, \vec{A}, \vec{B}}(V) = 1$  if for any  $\ell \in [c]$  the call  $\text{Witness}(V, I^*, A_\ell, B_\ell, u^*(I^*, A_\ell))$  returns a solutions  $x'_\ell$  such that  $B_V(x'_\ell - u^*(I^*, A_\ell)) = B_\ell$ , and  $\chi_{I^*, \vec{A}, \vec{B}}(V) = 0$  otherwise. The vector  $u^* = u^*(I^*, A_\ell)$  is the one defined in Equation 1.

**Corollary 18.** Assume that  $\text{OK}(V)$  is true. Then, the  $c^{\text{th}}$  power of the number  $\text{PO}(V)$  of Pareto-optimal solutions is at most  $\sum_{(I^*, \vec{A}) \in \mathcal{C}_c} \sum_{\vec{B} \in \mathbb{B}_\varepsilon^c} \chi_{I^*, \vec{A}, \vec{B}}(V)$ .

*Proof.* The  $c^{\text{th}}$  power of the number  $\text{PO}(V)$  of Pareto-optimal solutions equals the number of  $c$ -tuples  $(x_1, \dots, x_c)$  of Pareto-optimal solutions. Let  $(x_1, \dots, x_c)$  be such a  $c$ -tuple, let  $(I^*, \vec{A})$  be the  $V$ -certificate of  $(x_1, \dots, x_c)$ , and let  $B_\ell = B_V(x_\ell - u^*(I^*, A_\ell)) \in \mathbb{B}_\varepsilon$ . Due to Lemma 9,  $\text{Witness}(V, I^*, A_\ell, B_\ell, u^*(I^*, A_\ell))$  returns vector  $x_\ell$  for any  $\ell \in [c]$ . Hence,  $\chi_{I^*, \vec{A}, \vec{B}}(V) = 1$  for  $\vec{B} = (B_1, \dots, B_c)$ . To conclude the proof we show that the function  $(x_1, \dots, x_c) \mapsto (I^*, \vec{A}, \vec{B})$  is injective. Let  $(x_1, \dots, x_c)$  and  $(y_1, \dots, y_c)$  be distinct  $c$ -tuples of Pareto-optimal solutions, i.e., there is an index  $\ell \in [c]$  such that  $x_\ell \neq y_\ell$ , and let  $(I_1^*, \vec{A}^{(1)})$  and  $(I_2^*, \vec{A}^{(2)})$  be their  $V$ -certificates. If  $(I_1^*, \vec{A}^{(1)}) \neq (I_2^*, \vec{A}^{(2)})$ , then both tuples are mapped to distinct triplets. Otherwise,  $u^*(I_1^*, A_\ell^{(1)}) = u^*(I_2^*, A_\ell^{(2)})$  and thus  $B_V(x_\ell - u^*(I_1^*, A_\ell^{(1)})) \neq B_V(y_\ell - u^*(I_2^*, A_\ell^{(2)}))$  since  $\text{OK}(V)$  holds. Consequently, also in this case  $(x_1, \dots, x_c)$  and  $(y_1, \dots, y_c)$  are mapped to distinct triplets.  $\square$

Corollary 18 immediately implies a bound on the  $c^{\text{th}}$  moment of the number of Pareto-optimal solutions.

**Corollary 19.** The  $c^{\text{th}}$  moment of the number of Pareto-optimal solutions is bounded by

$$\mathbf{E}_V[\text{PO}^c(V)] \leq \sum_{(I^*, \vec{A}) \in \mathcal{C}_c} \sum_{\vec{B} \in \mathbb{B}_\varepsilon^c} \mathbf{Pr}_V \left[ E_{I^*, \vec{A}, \vec{B}} \right] + 2^n \cdot \mathbf{Pr}_V \left[ \overline{\text{OK}(V)} \right]$$

where  $E_{I^*, \vec{A}, \vec{B}}$  denotes the event that  $\chi_{I^*, \vec{A}, \vec{B}}(V) = 1$ .

We omit the proof since it is exactly the same as the one of Corollary 11.

**Lemma 20.** The size of the restricted certificate space is bounded by  $|\mathcal{C}_c| \leq 2^{c^2 \cdot (d+1)^2} \cdot n^{cd} = O(n^{cd})$  for fixed  $c$  and  $d$ .

*Proof.* Let  $(I^*, \vec{A})$  be an arbitrary  $c$ -certificate. Each matrix  $A_\ell$  is a  $|I^*| \times (d+1)$ -matrix. The tuple  $I^*$  can be written as  $I^* = (i_d^{(1)}, \dots, i_1^{(1)}, i^{*(1)}, \dots, i_d^{(c)}, \dots, i_1^{(c)}, i^{*(c)})$ , created by  $c$  successive calls of the **Witness** function, where the indices  $i^{*(\ell)}$  are chosen deterministically in Definition 7. Since  $|I^*| = c \cdot (d+1)$  the claim follows.  $\square$

**Corollary 21.** For an arbitrary restricted  $c$ -certificate  $(I^*, \vec{A})$  and an arbitrary vector  $\vec{B} \in \mathbb{B}_\varepsilon^c$  of  $\varepsilon$ -boxes the probability of the event  $E_{I^*, \vec{A}, \vec{B}}$  is bounded by

$$\Pr_V \left[ E_{I^*, \vec{A}, \vec{B}} \right] \leq (2\gamma)^{\gamma - cd} \cdot \phi^\gamma \cdot \varepsilon^{cd}$$

and by

$$\Pr_V \left[ E_{I^*, \vec{A}, \vec{B}} \right] \leq 2^{cd} \cdot \gamma^{\gamma - cd} \cdot \phi^{cd} \cdot \varepsilon^{cd}$$

if all densities are quasiconcave, where  $\gamma = cd \cdot (d + 1)$ .

*Proof.* For  $k \in [d]$  and  $\ell \in [c]$  consider the matrices  $Q_k(I^*, A_\ell, u_\ell^*)$  for  $u_\ell^* = u^*(I^*, A_\ell)$  defined in Equation (2). Due to Lemma 13 the output of the call  $\text{Witness}(V, I^*, A_\ell, B_\ell, u_\ell^*)$  is determined if  $V_{I^*}$  and the linear combinations  $V_{I^*}^k \cdot q$  for any index  $k \in [d]$  and any column  $q$  of the matrix  $Q_k^{(\ell)} = Q_k(I^*, A_\ell, u_\ell^*)$  are given. With the same argument as in the proof of Corollary 21 event  $E_{I^*, \vec{A}, \vec{B}}$  occurs if and only if  $V_{I^*} \cdot [p^{(\ell, 1)}, \dots, p^{(\ell, d)}]$  falls into some  $d$ -dimensional hypercube  $C_\ell$  with side length  $\varepsilon$  depending on the linear combinations  $V_{I^*} \cdot Q_k^{(\ell)}$ . In this notation,  $p^{(\ell, t)}$  is short for  $p^{(t)}(I^*, A_\ell, u_\ell^*)$ .

Now, consider the matrix

$$Q'_k = \left[ Q_k^{(1)}, p^{(1, k)}, \dots, Q_k^{(c)}, p^{(c, k)} \right].$$

Due to Lemma 14 this is a lower block triangular matrix, due to Lemma 15 the columns of  $[Q_k^{(\ell)}, p^{(\ell, k)}]$  are linearly independent. Hence, matrix  $Q'_k$  is an invertible matrix and the same holds for the block diagonal matrix

$$Q' = \begin{bmatrix} Q'_1 & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O} \\ \mathbb{O} & \dots & \mathbb{O} & Q'_d \end{bmatrix} \in \{-1, 0, 1\}^{(d \cdot |I^*|)^2}.$$

We permute the columns of  $Q'$  to obtain a matrix  $Q$  where the last  $c \cdot d$  columns belong to the columns  $p^{(1, 1)}, \dots, p^{(1, d)}, \dots, p^{(c, 1)}, \dots, p^{(c, d)}$ . We assume the rows of  $Q$  to be labeled by  $Q_{j_1, 1}, \dots, Q_{j_m, 1}, \dots, Q_{j_1, d}, \dots, Q_{j_m, d}$  where  $I^* = (j_1, \dots, j_m)$  and introduce random variables  $X_{j, k} = V_j^k$ ,  $j \in I^*$ ,  $k \in [d]$ , indexed the same way as the rows of  $Q$ . Event  $E_{I^*, \vec{A}, \vec{B}}$  holds if and only if the  $cd$  linear combinations of the variables  $X_{j, k}$  given by the last  $cd$  columns of  $Q$  fall into the  $cd$ -dimensional hypercube  $C = \prod_{\ell=1}^c C_\ell$  with side length  $\varepsilon$  depending on the linear combinations of the variables  $X_{j, k}$  given by the remaining columns of  $Q$ . The claim follows by applying Theorem 35 for the matrix  $Q^T$  and due to the fact that the number of rows of  $Q$  is  $d \cdot |I^*| = cd \cdot (d + 1) = \gamma$ .  $\square$

*Proof of Theorem 4.* In the proof of Theorem 3 we show that the probability that the OK-event does not hold is bounded by  $2^{2n+1}d\phi\varepsilon$ . For  $\gamma = cd \cdot (d + 1)$ , we set  $s = (2\gamma)^{\gamma - cd} \cdot \phi^\gamma$  for general densities and  $s = 2^{cd} \cdot \gamma^{\gamma - cd} \cdot \phi^{cd}$  in the case of quasiconcave density functions. Then, we obtain

$$\begin{aligned} \mathbf{E}_V[\text{PO}^c(V)] &\leq \sum_{(I^*, \vec{A}) \in \mathcal{C}_c} \sum_{\vec{B} \in \mathbb{B}_\varepsilon^c} \Pr_V \left[ E_{I^*, \vec{A}, \vec{B}} \right] + 2^n \cdot \Pr_V \left[ \overline{\text{OK}}(V) \right] \\ &\leq \sum_{(I^*, \vec{A}) \in \mathcal{C}_c} \sum_{\vec{B} \in \mathbb{B}_\varepsilon^c} s \cdot \varepsilon^{cd} + 2^n \cdot 2^{2n+1}d\phi\varepsilon \\ &\leq 2^{c^2 \cdot (d+1)^2} \cdot n^{cd} \cdot \left( \left( \frac{2n}{\varepsilon} \right)^d \right)^c \cdot s \cdot \varepsilon^{cd} + 2^{3n+1}d\phi\varepsilon \end{aligned}$$

$$\leq 2^{c^2 \cdot (d+1)^2 + cd} \cdot n^{2cd} \cdot s + 2^{3n+1} d \phi \varepsilon.$$

The first inequality is due to Corollary 19. The second inequality is due to Corollary 21. The third inequality stems from Lemma 20. Since this bound is true for arbitrarily small  $\varepsilon > 0$ , the correctness of Theorem 4 follows.  $\square$

The proof of Theorem 4 yields that

$$\mathbf{E}_V[\text{PO}^c(V)] \leq s_c = 4^{c^2(d+1)^2} \cdot (cd(d+1))^{cd^2} \cdot n^{2cd} \cdot \phi^{c\beta}$$

where  $\beta = d(d+1)$  in general and  $\beta = d$  for quasiconcave densities. With the following Corollary we bound the probability that  $\text{PO}(V)$  exceeds a certain multiple of  $s_1$ . We obtain a significantly better concentration bound than the one we would obtain by applying Markov's Inequality for the first moment.

**Corollary 22.** *The probability that the number of Pareto-optimal solutions is at least  $k \cdot s_1$  is bounded by*

$$\Pr_V[\text{PO}(V) \geq k \cdot s_1] \leq \left(\frac{1}{k}\right)^{\frac{1}{2} \cdot \left\lfloor \frac{\log_8 k}{2(d+1)^2} \right\rfloor}.$$

*Proof.* We bound the probability as follows:

$$\begin{aligned} \Pr_V[\text{PO}(V) \geq k \cdot s_1] &= \Pr_V[\text{PO}^c(V) \geq k^c \cdot s_1^c] = \Pr_V\left[\text{PO}^c(V) \geq \frac{k^c \cdot s_1^c}{\mathbf{E}_V[\text{PO}^c(V)]} \cdot \mathbf{E}_V[\text{PO}^c(V)]\right] \\ &\leq \Pr_V\left[\text{PO}^c(V) \geq \frac{k^c \cdot s_1^c}{s_c} \cdot \mathbf{E}_V[\text{PO}^c(V)]\right] \leq \frac{s_c}{k^c \cdot s_1^c} \\ &= \frac{4^{c^2(d+1)^2} \cdot (cd(d+1))^{cd^2} \cdot n^{2cd} \cdot \phi^{c\beta}}{k^c \cdot 4^{c(d+1)^2} \cdot (d(d+1))^{cd^2} \cdot n^{2cd} \cdot \phi^{c\beta}} \leq \frac{4^{c^2(d+1)^2} \cdot c^{cd^2}}{k^c} \\ &= \left(\frac{4^{c(d+1)^2} \cdot 4^{d^2 \cdot \log_4 c}}{k}\right)^c \leq \left(\frac{8^{c(d+1)^2}}{k}\right)^c. \end{aligned}$$

We obtain the claimed bound by setting  $c = \lfloor c^* \rfloor$  where  $c^*$  is chosen such that  $8^{c^*(d+1)^2} = k^{1/2}$ , i.e.,  $c^* = (\log_8 k)/(2(d+1)^2)$ .  $\square$

## 5 Zero-preserving Perturbations

To prove Theorem 1 we will first show that we can concentrate on a special class of instances.

**Lemma 23.** *Without loss of generality in each objective function except for the adversarial one there are more than  $f(d)$  perturbed coefficients, i.e., coefficients that are not deterministically set to zero, where  $f$  is an arbitrary fixed function.*

*Proof.* For  $k \in [d]$  let  $P_k$  be the tuple of indices  $i$  such that  $V_i^k$  is a perturbed coefficient. Let  $K$  be the tuple of indices  $k$  for which  $|P_k| \leq f(d)$ , let  $P = \bigcup_{k \in K} P_k$ , and consider the decomposition of  $\mathcal{S}$  into subsets of solutions  $\mathcal{S}_v = \{x \in \mathcal{S} : x|_P = v\}$ ,  $v \in \{0, 1\}^{|P|}$ . Let  $x \in \mathcal{S}_v$  be an arbitrary solution. If  $x$  is Pareto-optimal with respect to  $\mathcal{S}$  and  $V^{1 \dots d+1}$ , then  $x$  is also Pareto-optimal with respect to  $\mathcal{S}_v$  and the objective functions  $V^k$ ,  $k \in [d+1] \setminus K$ . As all remaining objective functions  $V^k$ ,  $k \in [d+1] \setminus K$ , have more than  $f(d)$  perturbed coefficients, the instance with these objective functions and  $\mathcal{S}_v$  as set of feasible solutions has the desired form. Since there are  $2^{|P|} \leq 2^{d \cdot f(d)} = O(1)$  choices for  $v$ , the smoothed number of Pareto-optimal solutions for general  $\phi$ -smooth instances is only by a constant factor larger than the smoothed number of Pareto-optimal solutions for  $\phi$ -smooth instances of the form mentioned in the lemma.  $\square$

**Lemma 24.** *Without loss of generality for every  $i \in [n]$  exactly one of the coefficients  $V_i^1, \dots, V_i^d$  is perturbed, whereas the others are deterministically set to zero.*

*Proof.* We first show how to decrease the number of indices  $i$  for which  $V_i^1, \dots, V_i^d$  is perturbed to at most one. For this, let  $\mathcal{S}' = \{(x, x, \dots, x) \mid x \in \mathcal{S}\} \subseteq \{0, 1\}^{dn}$  be the set of feasible solutions that contains for every  $x \in \mathcal{S}$  the solution  $x^d \in \{0, 1\}^{dn}$  that consists of  $d$  copies of  $x$ . For  $k \in [d]$  we define a linear objective function  $W^k: \mathcal{S}' \rightarrow \mathbb{R}$  in which all coefficients  $W_i^k$  with  $i \notin \{(k-1)n+1, \dots, kn\}$  are deterministically set to zero. The coefficients  $W_{(k-1)n+1}^k, \dots, W_{kn}^k$  are chosen as the coefficients  $V_1^k, \dots, V_n^k$ , i.e., either randomly according to a density  $f_i^k$  or zero deterministically. The objective function  $W^{d+1}$  maps every solution  $x^d \in \mathcal{S}'$  to  $V^{d+1}(x)$ . The instance consisting of  $\mathcal{S}'$  and the objective functions  $W^1, \dots, W^{d+1}$  has the desired property that every binary variable appears in at most one of the objective functions  $W^1, \dots, W^d$  and it has the same smoothed number of Pareto-optimal solutions as the instance consisting of  $\mathcal{S}$  and the objective functions  $V^1, \dots, V^{d+1}$ . For every  $i \in [dn]$  for which none of the coefficients  $W_i^1, \dots, W_i^d$  is perturbed we can eliminate the corresponding binary variable from  $\mathcal{S}'$ .

This shows that any  $\phi$ -smooth instance with  $\mathcal{S} \subseteq \{0, 1\}^n$  can be transformed into another  $\phi$ -smooth instance with  $\mathcal{S} \subseteq \{0, 1\}^\ell$  with  $\ell \leq dn$  in which every binary variable appears in exactly one objective function and that has the same smoothed number of Pareto-optimal solutions. As the bound proven in Theorem 1 depends polynomially on the number of binary variables, we lose only a constant factor by going from  $\mathcal{S} \subseteq \{0, 1\}^n$  to  $\mathcal{S}' \subseteq \{0, 1\}^{dn}$ .  $\square$

In the remainder of this chapter we focus on instances having the structure described in Lemma 23 and Lemma 24. Then,  $(P_1, \dots, P_d)$  is a partition of  $[n]$ , where  $P_t$  denotes the tuple of indices  $i$  for which  $V_i^t$  is perturbed. This structure also ensures that  $B_V(x) \in \mathbb{B}_\varepsilon$  holds for any vector  $\{-1, 0, 1\}^n$ .

We consider the following variant of the **Witness** function, which gets as parameters besides the usual  $V$  and  $x$ , a set  $K \subseteq [d]$  of objective functions, a call number  $r \in \mathbb{N}$ , and a set  $I \subseteq [n]$  of indices. In a call of the **Witness<sub>0</sub>** function only the adversarial objective function  $V^{d+1}$  and the objective functions  $V^t$  with  $t \in K$  are considered, and the solution set is restricted to solutions that agree with  $x$  in all positions  $P_k$  with  $k \notin K$ . Additionally, as in the **Witness** function for higher moments, only solutions are considered that agree with  $x$  in all positions  $i \in I$ . By the right choice of  $I$ , we can avoid choosing an index multiple times in different calls of the **Witness** function. The parameter  $r$  simply corresponds to the number of the current call of the **Witness<sub>0</sub>** function. The **Witness<sub>0</sub>** function always returns some subset of  $\mathcal{S}$ .

**Witness<sub>0</sub>**( $V, x, K, r, I$ )

- 1: Let  $K$  be of the form  $K = (k_1, \dots, k_{d'})$ .
- 2: Set  $k_{d'+1} = d + 1$ .
- 3: Set  $\mathcal{R}_{d'+1}^{(r)} = \mathcal{S} \cap \mathcal{S}_I(x) \cap \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$ .
- 4: **if**  $d' = 0$  **then return**  $\mathcal{R}_{d'+1}^{(r)}$
- 5: **for**  $t = d', d' - 1, \dots, 0$  **do**
- 6:   Set  $\mathcal{C}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} : V^{k_1 \dots k_t} z < V^{k_1 \dots k_t} x\}$ .
- 7:   **if**  $\mathcal{C}_t^{(r)} \neq \emptyset$  **then**
- 8:     Set  $X_t^{(r)} = \arg \min \{V^{k_{t+1}} z : z \in \mathcal{C}_t^{(r)}\}$ .
- 9:     Let  $x^{(r,t)}$  be the canonically first vector in  $X_t^{(r)}$ .
- 10:    Let  $K_{\text{EQ}} \subseteq K$  be the tuple of indices  $k$  such that  $x^{(r,t)}|_{P_k} = x|_{P_k}$ .
- 11:    Let  $K_{\text{NEQ}} \subseteq K$  be the tuple of the remaining indices.
- 12:    **for**  $k \in K$  **do**
- 13:     **if**  $k \in K_{\text{EQ}}$  **then**
- 14:       Set  $r_k = r$ .

```

15:     else
16:         Determine the first index  $i \in P_k$  such that  $x_i^{(r,t)} \neq x_i$ .
17:          $I \leftarrow I \cup \{i\}$ 
18:     end if
19: end for
20: if  $K_{\text{EQ}} = ()$  then
21:     Set  $\mathcal{R}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} \cap \mathcal{S}_I(x) : V^{k_{t+1}}z < V^{k_{t+1}}x^{(r,t)}\}$ .
22: else
23:     Set  $t_r = t$ .
24:     return  $\text{Witness}_0(V, x, K_{\text{NEQ}}, r+1, I)$ 
25: end if
26: else
27:     for  $k \in K$  do
28:         Set  $i_k = \min(P_k \setminus I)$ .
29:          $I \leftarrow I \cup \{i_k\}$ .
30:     end for
31:     Set  $x_i^{(r,t)} = \begin{cases} \bar{x}_i & : i \in \{i_{k_1}, \dots, i_{k_{d'}}\}, \\ x_i & : \text{otherwise.} \end{cases}$ 
32:     Set  $\mathcal{R}_t^{(r)} = \mathcal{R}_{t+1}^{(r)} \cap \mathcal{S}_I(x)$ .
33: end if
34: end for
35: return  $\emptyset$ 

```

The index  $r_k$  defined in line 14 is the number of the last call in which the objective function  $V^k$  has been considered. In the following we use the term *round* to denote an iteration of the for-loop starting in line 5. The index  $t_r$  defined in line 23 is the number of the round in call number  $r$  of  $\text{Witness}_0$  in which the recursive call of  $\text{Witness}_0$  was made. In line 6 we define the winner set  $\mathcal{C}_t^{(r)}$  using the functions  $V^{k_1}, \dots, V^{k_t}$ . Note that in round  $t = 0$  there is no restriction and this definition simplifies to  $\mathcal{C}_0^{(r)} = \mathcal{R}_1^{(r)}$ . In line 16 it is always possible to find an index from  $i \in P_k$  on which the current vector  $x^{(r,t)}$  and  $x$  disagree because this line is only reached for  $k \in K_{\text{NEQ}}$ , i.e., only if  $x^{(r,t)}|_{P_k} \neq x|_{P_k}$ . In order for line 28 to be feasible, we have to guarantee that  $P_k \setminus I \neq \emptyset$ . This follows if we assume  $|P_k| > d(d+1)$  (which is in accordance with Lemma 23 without loss of generality) because there are at most  $d$  calls of  $\text{Witness}_0$  with non-empty  $K$  with at most  $d+1$  rounds each and in each round at most one index from  $P_k$  is added to  $I$ .

One important property that we will exploit later is that every Pareto-optimal solution  $x$  is also Pareto-optimal with respect to the objective functions  $V^k$  with  $k \in K_{\text{NEQ}}$  if the set  $\mathcal{S}$  is restricted to solutions that agree with  $x$  in all positions  $i \in \bigcup_{k \in K_{\text{EQ}}} P_k$ , i.e., solutions  $z$  for which  $V^k z = V^k x$  for every  $k \in K_{\text{EQ}}$ . This property guarantees that whenever  $\text{Witness}_0$  is called with a Pareto-optimal solution  $x$  as parameter, also in the recursive call in line 24  $x$  is Pareto-optimal with respect to the restricted solution set and the remaining objective functions.

In the remainder of this section we only consider the case that  $x$  is Pareto-optimal. Unless stated otherwise, we consider the case that the  $\text{OK}_0$ -event occurs. That means that  $|V^k \cdot (y - z)| \geq \varepsilon$  for every  $k \in [d]$  and for any two vectors  $y, z \in \mathcal{S}$  for which  $y|_{P_k} \neq z|_{P_k}$ .

**Lemma 25.** *The call  $\text{Witness}_0(V, x, [d], 1, ())$  returns the set  $\{x\}$ . Moreover,  $x^{(r^*, t_{r^*})} = x$  for the last constructed vector  $x^{(r^*, t_{r^*})}$ , i.e., for  $r^* = \max\{r_1, \dots, r_d\}$ .*

*Proof.* We show that for all tuples  $K \subseteq [d]$  and  $I \subseteq [n]$  and any call number  $r$  the result of the call  $\text{Witness}_0(V, x, K, r, I)$  is the set  $\{x\}$  by induction over the size of  $K$ . If  $K$  is the empty tuple, then  $\mathcal{R}_{d'+1}^{(r)} = \mathcal{S} \cap \mathcal{S}_I(x) \cap \bigcap_{k \in [d]} \mathcal{S}_{P_k}(x)$ , i.e., the set  $\mathcal{R}_{d'+1}^{(r)}$ , which will be returned in line 4, consists of all solutions that are identical with  $x$  in every index. Consequently,  $\mathcal{R}_{d'+1}^{(r)} = \{x\}$  in that case.



Now we consider non-empty tuples  $K \subseteq [d]$ . By the induction hypothesis it suffices to show that there is a call to the **Witness**<sub>0</sub> function before the end of the loop. Hence, let us assume that there is no such call during the rounds  $t = d', d' - 1, \dots, 1$ . As for the simple variant of the **Witness** function we show two claims for that case:

**Claim 3.** *There is no  $t \in [d' + 1]$  such that there exists a solution  $z \in \mathcal{R}_t^{(r)}$  that dominates  $x$  with respect to the objective functions  $V^{k_1}, \dots, V^{k_t}$ .*

**Claim 4.** *For any  $t \in [d' + 1]$  solution  $x$  is an element of  $\mathcal{R}_t^{(r)}$ .*

As there is no call to the **Witness**<sub>0</sub> function during the first rounds we always get  $K_{\text{EQ}} = ()$  whenever  $\mathcal{C}_t^{(r)} \neq \emptyset$ . Hence, the rounds are basically the same as the rounds in the simple **Witness** function and both claims hold due to the same arguments used to prove Lemma 6 and the fact that  $x \in \mathcal{S}_J(x)$  for any index tuple  $J \subseteq [n]$ .

Using Claim 3 and Claim 4 we can prove the lemma: Solution  $x$  is an element of  $\mathcal{C}_0^{(r)} = \mathcal{R}_1^{(r)}$  according to Claim 4. In particular,  $\mathcal{C}_0^{(r)} \neq \emptyset$ . Due to Claim 3 we get  $x \in X_0^{(r)}$  and, hence,  $V^{k_1}x^{(r,0)} = V^{k_1}x$ . As the  $\text{OK}_0$ -event holds, this implies  $x^{(r,0)}|_{P_{k_1}} = x|_{P_{k_1}}$ , leading to  $k_1 \in K_{\text{EQ}}$ . Consequently, the tuple  $K_{\text{EQ}}$  is not empty and we call the **Witness**<sub>0</sub> function again.

To conclude the proof of the lemma we have to show the equality  $x^{(r^*, t_{r^*})} = x$ . In the call of the **Witness**<sub>0</sub> function with number  $r^*$  the last recursive call in line 24 is made. We know that  $x^{(r^*, t_{r^*})}|_{P_k} = x|_{P_k}$  for any  $k \in [d] \setminus K$  where  $K$  is the parameter of call number  $r^*$  because  $x^{(r^*, t_{r^*})} \in \mathcal{R}_{t_{r^*}+1}^{(r^*)} \subseteq \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$ . Additionally, since the recursive call that is made in line 24 terminates immediately, all objective functions are removed from consideration in the call with number  $r^*$ , i.e.,  $K_{\text{EQ}} = K$  in this call. Hence, the equation  $x^{(r^*, t_{r^*})}|_{P_k} = x|_{P_k}$  holds also for any  $k \in K$ . Together this implies  $x^{(r^*, t_{r^*})} = x$ .  $\square$

Like for the simple **Witness** function, we show that it is enough to know some information about one run of the **Witness**<sub>0</sub> function to reconstruct the vector  $x$ . As before, we call this data the certificate of  $x$ .

**Definition 26.** *Let  $r_1, \dots, r_d$  and  $t_1, \dots, t_{r^*}$ , where  $r^* = \max\{r_1, \dots, r_d\}$ , be the indices and let  $x^{(r,t)}$  be the vectors constructed during the execution of the call **Witness**<sub>0</sub>( $V, x, [d], 1, ()$ ) and let  $d'_r$  be the value of  $d'$  in the  $r^{\text{th}}$  call of the **Witness**<sub>0</sub> function. Furthermore, consider the tuple  $I$  at the moment when the last call terminates. The pair  $(I^*, A)$ , where  $I^* = I \cup (i_1^*, \dots, i_d^*)$ ,  $i_k^* = \min(P_K \setminus I)$ , and  $A = \left[ x^{(1,d'_1)}, \dots, x^{(1,t_1)}, \dots, x^{(r^*, d'_{r^*})}, \dots, x^{(r^*, t_{r^*})} \right]_{I^*}$ , is called the  $V$ -certificate of  $x$ . We label the columns of  $A$  by  $a^{(r,t)}$ . Moreover, we call a pair  $(I', A')$  a certificate if there is some realization  $V$  such that  $\text{OK}_0(V)$  is true and there exists a Pareto-optimal  $x \in \mathcal{S}$  such that  $(I', A')$  is the  $V$ -certificate of  $x$ . By  $\mathcal{C}$  we denote the set of all certificates.*

We assume that the indices  $r_k$  and  $t_r$  (and hence also the indices  $d'_r$ ) are implicitly encoded in a given certificate. Later we will take these indices into consideration again when we count the number of possible certificates.

**Lemma 27.** *Let  $V$  be a realization for which  $\text{OK}_0(V)$  is true and let  $(I^*, A)$  be a  $V$ -certificate of some Pareto-optimal  $x$ . Let  $A$  be of the form  $A = \left[ a^{(1,d'_1)}, \dots, a^{(1,t_1)}, \dots, a^{(r^*, d'_{r^*})}, \dots, a^{(r^*, t_{r^*})} \right]$ . For fixed  $k \in [d]$  set  $M = \left[ a^{(1,d'_1)}, \dots, a^{(1,t_1)}, \dots, a^{(r_k, d'_{r_k})}, \dots, a^{(r_k, t_{r_k})} \right]_J$ , where  $J = I^* \cap P_k =:$*

$(j_1, \dots, j_m)$ , and set  $y = x|_J$ . Then,  $M$  is of the form

$$M = \begin{bmatrix} \overline{y_{j_1}} & y_{j_1} & \dots & y_{j_1} \\ * & \ddots & \ddots & \vdots \\ \vdots & \ddots & \overline{y_{j_{m-1}}} & y_{j_{m-1}} \\ * & \dots & * & y_{j_m} \end{bmatrix} \in \{0, 1\}^{|J| \times |J|},$$

where each ‘ $*$ ’ can either be 0 or 1 independently of the other ‘ $*$ ’-entries.

*Proof.* Consider the call  $\text{Witness}_0(V, x, [d], 1, ())$  and all subsequent calls  $\text{Witness}_0(V, x, K, r, I)$ . By definition of  $r_k$  we have  $r \leq r_k \iff k \in K$ . In each call where  $r \leq r_k$  one vector  $x^{(r,t)}$  is constructed each round. Also, in each round except for the last round of the  $r_k^{\text{th}}$  call one index  $i \in P_k$  is chosen and added to  $I$ . Since  $J$  consists of the chosen indices  $i \in P_k$  and the additional index  $i_k^*$ , matrix  $M$  is a square matrix.

We first consider the last column of  $M$ . As  $x^{(r_k, t_{r_k})}$  is the last vector constructed before  $k$  is removed from  $K$ , index  $k$  must be an element of  $K_{\text{EQ}}$  at that time, i.e.,  $x^{(r_k, t_{r_k})}|_{P_k} = x|_{P_k}$ . Hence, the last column of  $M$  has the claimed form.

Now consider the remaining columns of  $M$ . Due to the construction of the set  $\mathcal{R}_t^{(r)}$  in line 21 or in line 32, all vectors  $z$  in a set  $\mathcal{R}_t^{(r)}$  agree with  $x$  in the previously chosen indices  $i$ . As in the case  $\mathcal{C}_t^{(r)} \neq \emptyset$  vector  $x^{(r,t)}$  is an element of  $\mathcal{R}_{t+1}^{(r)}$  and in the case  $\mathcal{C}_t^{(r)} = \emptyset$  vector  $x^{(r,t)}$  is constructed appropriately, the upper triangle of  $M$ , excluding the principal diagonal, has the claimed form. The form of the principal diagonal follows from the choice of index  $i \in P_k$  such that  $x_i^{(r,t)} \neq x_i$  in line 16 or from the construction of  $x^{(r,t)}$  in line 31.  $\square$

As in the simple case we now consider the following  $\text{Witness}_0$  function that is based on the guess of the  $V$ -certificate of  $x$ , a shift vector  $u \in \{0, 1\}^n$  and the  $\varepsilon$ -box  $B = B_V(x - u)$ .

$\text{Witness}_0(V, K, r, I^*, A, \mathcal{S}', B, u)$

- 1: Let  $K$  be of the form  $K = (k_1, \dots, k_{d'})$ .
- 2: Set  $k_{d'+1} = d + 1$ .
- 3: Let  $b$  be the corner of  $B$ .
- 4: Set  $\mathcal{R}_{d'+1}^{(r)} = \mathcal{S}' \cap \bigcup_{t'=t_r}^{d'} \mathcal{S}_{I^*}(a^{(r,t')})$ .
- 5: **if**  $d' = 0$  **then return**  $\mathcal{R}_{d'+1}^{(r)}$
- 6: **for**  $t = d', d' - 1, \dots, 0$  **do**
- 7:   Set  $\mathcal{C}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} \cap \mathcal{S}_{I^*}(a^{(r,t)}) : V^{k_1 \dots k_t} \cdot (z - u) \leq b|_{k_1 \dots k_t}\}$ .
- 8:   **if**  $\mathcal{C}_t^{(r)} \neq \emptyset$  **then**
- 9:     Set  $X_t^{(r)} = \arg \min \{V^{k_{t+1}} z : z \in \mathcal{C}_t^{(r)}\}$ .
- 10:   Let  $x^{(r,t)}$  be the canonically first vector in  $X_t^{(r)}$ .
- 11:   **if**  $t = t_r$  **then**
- 12:     Let  $K_{\text{EQ}} \subseteq K$  be the tuple of indices  $k$  such that  $r_k = r$ .
- 13:     Let  $K_{\text{NEQ}} \subseteq K$  be the tuple of the remaining indices.
- 14:     **return**  $\text{Witness}_0(V, K_{\text{NEQ}}, r + 1, I^*, A, \mathcal{S}' \cap \bigcap_{k \in K_{\text{EQ}}} \mathcal{S}_{P_k}(x^{(r,t)}), B, u)$
- 15:   **else**
- 16:     Set  $\mathcal{R}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} \cap \bigcup_{t'=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,t')}) : V^{k_{t+1}} z < V^{k_{t+1}} x^{(r,t)}\}$ .
- 17:   **end if**
- 18:   **else**
- 19:     Set  $x^{(r,t)} = (\perp, \dots, \perp)$ .
- 20:     Set  $\mathcal{R}_t^{(r)} = \mathcal{R}_{t+1}^{(r)} \cap \bigcup_{t'=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,t')})$ .
- 21:   **end if**
- 22: **end for**

23: **return**  $\emptyset$

**Lemma 28.** *Let  $(I^*, A)$  be the  $V$ -certificate of  $x$ , let  $u \in \{0, 1\}^n$  be an arbitrary vector, and let  $B = B_V(x - u)$ . Then, the call  $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$  returns  $\{x\}$ .*

Before we give a formal proof of Lemma 28 we try to give some intuition for it. As for the simple variant of the **Witness** function we restrict the set of solutions to vectors that look like the vectors we want to reconstruct in the next rounds of the current call, i.e., we intersect the current set with the set  $\bigcup_{t'=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,t')})$  in round  $t$ . That way we only deal with subsets of the original sets, but we do not lose the vectors we want to reconstruct. In order to reconstruct the vectors, we need more information than in the simple variant: we need to know in which rounds the recursive calls of  $\text{Witness}_0$  are made, in each call we need to know which objective functions  $V^k$  must not be considered anymore, and for each of these objective functions we need to know the vector  $x|_{P_k}$ . The information when the recursive calls are made and which objective functions must not be considered anymore is given in the certificate: The variable  $t_r$  contains the round number when the recursive call is made. The index  $r_k$  contains the number of the call where index  $k$  has to be removed from  $K$ . Hence, index  $k$  is removed in the  $t_{r_k}^{\text{th}}$  round of call  $r_k$ . If we can reconstruct  $K_{\text{EQ}}$  and the vector  $x^{(r,t)}$  in the round where we make the recursive call, then we can also reconstruct the bits of  $x$  at indices  $i \in P_k$  for any index  $k \in K_{\text{EQ}}$  since  $x|_{P_k} = x^{(r,t)}|_{P_k}$  for these indices  $k$ .

*Proof of 28.* To prove the lemma we show that in the execution of  $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$  essentially the same sequence of vectors  $x^{(r,t)}$  as in the execution of  $\text{Witness}_0(V, x, [d], 1, ())$  is generated:

**Claim 5.** *Let  $\text{Witness}_0(V, x, K, r, I)$  be the  $r^{\text{th}}$  call to the  $\text{Witness}_0$  function resulting from the first call  $\text{Witness}_0(V, x, [d], 1, ())$ . Consider the call  $\text{Witness}_0(V, K, r, I^*, A, \mathcal{S}', B, u)$  where  $\mathcal{S}' = \mathcal{S} \cap \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$  and  $B = B_V(x - u)$ . If  $K = ()$ , then both calls return  $\{x\}$ . Otherwise, the following statements hold, where  $\mathcal{R}_t^{(r)}$ ,  $\mathcal{C}_t^{(r)}$ ,  $X_t^{(r)}$ ,  $x^{(r,t)}$ ,  $K'_{\text{EQ}}$ , and  $K'_{\text{NEQ}}$  denote the variables from the call  $\text{Witness}_0(V, K, r, I^*, A, \mathcal{S}', B, u)$ :*

- $\mathcal{R}_t^{(r)} \subseteq \mathcal{R}_t^{(r)}$  for any  $t \in \{t_r + 1, \dots, d' + 1\}$ ,
- $x^{(r,t)} = x^{(r,t)}$  for any  $t \in \{t_r, \dots, d'\}$  for which  $\mathcal{C}_t^{(r)} \neq \emptyset$ ,
- $x^{(r,t')} \in \mathcal{R}_t^{(r)}$  for any  $t \in \{t_r + 1, \dots, d' + 1\}$  and any  $t' \in \{t_r, \dots, t - 1\}$  for which  $\mathcal{C}_t^{(r)} \neq \emptyset$ , and
- the next calls are both in round  $t_r$  and are of the form  $\text{Witness}_0(V, x, K_1, r + 1, I')$  and  $\text{Witness}_0(V, K_2, r + 1, I^*, A, \mathcal{S}'', B, u)$  where  $\mathcal{S}'' = \mathcal{S} \cap \bigcap_{k \in [d] \setminus K_2} \mathcal{S}_{P_k}(x)$  and  $K_1 = K_2$ .

Using Claim 5, Lemma 28 follows immediately by an inductive argument. For  $K = ()$  Claim 5 is trivially true. To show the claim for  $K \neq ()$  we use induction over  $t$ .

First consider the case  $t = d' + 1$ , i.e., the moment before entering the loop. By assumption and the construction of the vectors  $a^{(r,t)}$  we have  $\mathcal{R}_{d'+1}^{(r)} = \mathcal{S}' \cap \bigcup_{t'=t_r}^{d'} \mathcal{S}_{I^*}(a^{(r,t')}) = (\mathcal{S} \cap \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)) \cap \bigcup_{t'=t_r}^{d'} \mathcal{S}_{I^*}(a^{(r,t')})$  and  $\mathcal{R}_{d'+1}^{(r)} = \mathcal{S} \cap \mathcal{S}_I(x) \cap \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$ , where  $I$  denotes the current index tuple. By construction, all vectors  $x^{(r,t')}$  agree with  $x$  in the indices  $i \in I$ , and since  $I \subseteq I^*$ , the relation  $\mathcal{R}_{d'+1}^{(r)} \subseteq \mathcal{R}_{d'+1}^{(r)}$  holds. As  $x^{(r,t')} \in \mathcal{R}_{d'+1}^{(r)} \subseteq \mathcal{S} \cap \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x) = \mathcal{S}'$  for any index  $t' \in \{t_r, \dots, d'\}$  where  $\mathcal{C}_{t'}^{(r)} \neq \emptyset$  and as  $x^{(r,t')} \in \mathcal{S}_{I^*}(a^{(r,t')})$ , also  $x^{(r,t')} \in \mathcal{R}_{d'+1}^{(r)}$  holds for these vectors. As we have not entered the loop yet, there is no recursive call for  $t = d' + 1$ .

Now consider the case  $t \leq d'$ . We have  $\mathcal{C}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} \cap \mathcal{S}_{I^*}(a^{(r,t)}) : V^{k_1 \dots k_t} \cdot (z - u) \leq b|_{k_1 \dots k_t}\}$  and  $\mathcal{C}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} : V^{k_1 \dots k_t} z < V^{k_1 \dots k_t} x\}$ . Since  $b$  is the corner of  $B = B_V(x - u)$  and since the  $\text{OK}_0$ -event occurs, the relations  $V^k \cdot (z - u) \leq b_k$  and  $V^k z < V^k x$  are equivalent

for any index  $k \in [d]$ . Consequently,  $\mathcal{C}_t^{(r)}$  is a subset of  $\mathcal{C}_t^{(r)}$  as  $\mathcal{R}_{t+1}'^{(r)}$  is a subset of  $\mathcal{R}_{t+1}^{(r)}$  by the induction hypothesis.

First, we analyze the case that  $\mathcal{C}_t^{(r)} \neq \emptyset$ . This is in particular the case for  $t = t_r$  due to the construction of  $t_r$ . Due to the induction hypothesis,  $x^{(r,t)}$  is an element of  $\mathcal{R}_{t+1}'^{(r)}$  and, due to the fact that  $x^{(r,t)} \in \mathcal{S}_{I^*}(a^{(r,t)})$ , an element of  $\mathcal{C}_t^{(r)}$ . Consequently,  $\mathcal{C}_t^{(r)} \neq \emptyset$  and, thus,  $x^{(r,t)} = x^{(r,t)}$ . If  $t > t_r$ , then in neither of the calls a recursive call is made because  $K_{\text{EQ}} = () = K'_{\text{EQ}}$  and the sets  $\mathcal{R}_t'^{(r)}$  and  $\mathcal{R}_t^{(r)}$  are defined as follows:  $\mathcal{R}_t'^{(r)} = \{z \in \mathcal{R}_{t+1}'^{(r)} \cap \bigcup_{t'=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,t')}) : V^{k_{t+1}}z < V^{k_{t+1}}x^{(r,t)}\}$  and  $\mathcal{R}_t^{(r)} = \{z \in \mathcal{R}_{t+1}^{(r)} \cap \mathcal{S}_I(x) : V^{k_{t+1}}z < V^{k_{t+1}}x^{(r,t)}\}$ . By the induction hypothesis and the construction of the vectors  $x^{(r,t')}$ ,  $\mathcal{R}_t'^{(r)}$  is a subset of  $\mathcal{R}_t^{(r)}$  and  $x^{(r,t')} \in \mathcal{R}_t'^{(r)}$  for any  $t' \in \{t_r, \dots, t-1\}$  where  $\mathcal{C}_{t'}^{(r)} \neq \emptyset$ . If  $t_r = t$ , then there is a call to  $\text{Witness}_0(V, x, K_{\text{NEQ}}, r+1, I)$  and a call to  $\text{Witness}_0(V, K'_{\text{NEQ}}, r+1, I^*, A, \mathcal{S}', B, u)$ , where  $\mathcal{S}' = \mathcal{S}' \cap \bigcap_{k \in K'_{\text{EQ}}} \mathcal{S}_{P_k}(x^{(r,t)})$ . In accordance with the definition of the indices  $r_k$  and the tuple  $K'_{\text{EQ}}$  we obtain the equivalence  $k \in K_{\text{EQ}} \iff r_k = r \iff k \in K'_{\text{EQ}}$  for any index  $k \in K$  and thus  $K'_{\text{EQ}} = K_{\text{EQ}}$  and  $K'_{\text{NEQ}} = K_{\text{NEQ}}$ . The definition of  $K_{\text{EQ}}$  yields  $x^{(r,t)}|_{P_k} = x|_{P_k}$ , i.e.,  $\bigcap_{k \in K_{\text{EQ}}} \mathcal{S}_{P_k}(x^{(r,t)}) = \bigcap_{k \in K_{\text{EQ}}} \mathcal{S}_{P_k}(x)$ . As  $\mathcal{S}' = \mathcal{S} \cap \bigcap_{k \in [d] \setminus K} \mathcal{S}_{P_k}(x)$  by assumption, we finally get  $\mathcal{S}' = \mathcal{S} \cap \bigcap_{k \in [d] \setminus K_{\text{NEQ}}} \mathcal{S}_{P_k}(x)$ .

Now, let us analyze the case  $\mathcal{C}_t^{(r)} = \emptyset$ , which implies  $\mathcal{C}_t'^{(r)} = \emptyset$ . Then, no recursive call is made and the sets  $\mathcal{R}_t'^{(r)}$  and  $\mathcal{R}_t^{(r)}$  are  $\mathcal{R}_t'^{(r)} = \mathcal{R}_{t+1}'^{(r)} \cap \bigcup_{t'=t_r}^{t-1} \mathcal{S}_{I^*}(a^{(r,t')})$  and  $\mathcal{R}_t^{(r)} = \mathcal{R}_{t+1}^{(r)} \cap \mathcal{S}_I(x)$ . The vectors  $x^{(r,t_r)}, \dots, x^{(r,t-1)}$  agree with  $x$  in the indices  $i \in I$ . Consequently,  $\mathcal{R}_t'^{(r)} \subseteq \mathcal{R}_t^{(r)}$  and  $x^{(r,t')} \in \mathcal{R}_t'^{(r)}$  for any  $t' \in \{t_r, \dots, t-1\}$  as  $x^{(r,t')} \in \mathcal{S}_{I^*}(a^{(r,t')})$  and as this relation holds for  $\mathcal{R}_{t+1}'^{(r)}$  due to the induction hypothesis. This concludes the proof.  $\square$

By the choice of the vector  $u$  we can control which information about  $V$  has to be known in order to be able to execute the call  $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$ . While Lemma 28 is correct for any choice of  $u \in \{0, 1\}^n$ , we have to choose  $u$  carefully in order for the following probabilistic analysis to work. Later we will see that  $u^* = u^*(I^*, A)$ ,

$$u_i^* = \begin{cases} \overline{x_i} & : i \in (i_1^*, \dots, i_d^*), \\ x_i & : i \in I^* \setminus (i_1^*, \dots, i_d^*), \\ 0 & : \text{otherwise,} \end{cases} \quad (3)$$

is well-suited for our purpose. Note, that  $i_k^* \in P_k$  are the indices that have been added to  $I$  in the definition of the  $V$ -certificate to obtain  $I^*$ . Furthermore,  $x_i$  is encoded in the last column of  $A$  for any index  $i \in I^*$ . Hence, vector  $u^*$  can be defined with the help of the  $V$ -certificate of  $x$ ; we do not have to know the vector  $x$  itself.

Next we bound the number of Pareto-optimal solutions. For this, consider the following functions  $\chi_{I^*, A, B}(V)$  parameterized by an arbitrary certificate  $(I^*, A) \in \mathcal{C}$  and an arbitrary  $\varepsilon$ -box  $B \in \mathbb{B}_\varepsilon$  and defined as follows:  $\chi_{I^*, A, B}(V) = 1$  if the call  $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u^*(I^*, A))$  returns a set  $\{x'\}$  such that  $B_V(x' - u^*(I^*, A)) = B$ , and  $\chi_{I^*, A, B}(V) = 0$  otherwise.

**Corollary 29.** *Assume that  $\text{OK}_0(V)$  is true. Then, the number  $\text{PO}(V)$  of Pareto-optimal solutions is at most  $\sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I^*, A, B}(V)$ .*

*Proof.* Let  $x$  be a Pareto-optimal solution, let  $(I^*, A)$  be the  $V$ -certificate of  $x$ , and let  $B = B_V(x - u^*(I^*, A)) \in \mathbb{B}_\varepsilon$ . Due to Lemma 28,  $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u^*(I^*, A))$  returns  $\{x\}$ . Hence,  $\chi_{I^*, A, B}(V) = 1$ . It remains to show that the function  $x \mapsto (I^*, A, B')$  is injective. Let  $x_1$  and  $x_2$  be distinct Pareto-optimal solutions and let  $(I_1^*, A_1)$  and  $(I_2^*, A_2)$  be the  $V$ -certificates of  $x_1$  and  $x_2$ , respectively. If  $(I_1^*, A_1) \neq (I_2^*, A_2)$ , then  $x_1$  and  $x_2$  are mapped to distinct triplets. Otherwise,  $u^*(I_1^*, A_1) = u^*(I_2^*, A_2)$  and, hence,  $B_V(x_1 - u^*(I_1^*, A_1)) \neq B_V(x_2 - u^*(I_2^*, A_2))$  because  $\text{OK}_0(V)$  is true. Consequently, also in this case  $x_1$  and  $x_2$  are mapped to distinct triplets.  $\square$

Corollary 29 immediately implies a bound on the expected number of Pareto-optimal solutions.

**Corollary 30.** *The expected number of Pareto-optimal solutions is bounded by*

$$\mathbf{E}_V[\text{PO}(V)] \leq \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \mathbf{Pr}_V[E_{I^*, A, B}] + 2^n \cdot \mathbf{Pr}_V[\overline{\text{OK}_0(V)}]$$

where  $E_{I^*, A, B}$  denotes the event that the call  $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u^*(I^*, A))$  returns a set  $\{x'\}$  such that  $B_V(x' - u^*(I^*, A)) = B$ .

*Proof.* Using Corollary 29 we obtain

$$\begin{aligned} \mathbf{E}_V[\text{PO}(V)] &= \mathbf{E}_V[\text{PO}(V) \mid \text{OK}_0(V)] \cdot \mathbf{Pr}_V[\text{OK}_0(V)] + \mathbf{E}_V[\text{PO}(V) \mid \overline{\text{OK}_0(V)}] \cdot \mathbf{Pr}_V[\overline{\text{OK}_0(V)}] \\ &\leq \mathbf{E}_V \left[ \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I^*, A, B}(V) \mid \text{OK}_0(V) \right] \cdot \mathbf{Pr}_V[\text{OK}_0(V)] + 2^n \cdot \mathbf{Pr}_V[\overline{\text{OK}_0(V)}] \\ &\leq \mathbf{E}_V \left[ \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \chi_{I^*, A, B}(V) \right] + 2^n \cdot \mathbf{Pr}_V[\overline{\text{OK}_0(V)}] \\ &= \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \mathbf{Pr}_V[E_{I^*, A, B}] + 2^n \cdot \mathbf{Pr}_V[\overline{\text{OK}_0(V)}]. \quad \square \end{aligned}$$

We will see that the first term of the sum in Corollary 30 can be bounded independently of  $\varepsilon$  and that the limit of the second term is 0 for  $\varepsilon \rightarrow 0$ . First of all, we analyze the size of the certificate space.

**Lemma 31.** *The size of the certificate space is bounded by  $|\mathcal{C}| \leq 2^{(d+1)^5} \cdot d^{2d+3} \cdot n^{d^2 \cdot (d+1)} = O(n^{d^2 \cdot (d+1)})$ .*

*Proof.* Consider the execution of the call  $\text{Witness}_0(V, x, [d], 1, ())$ . Including that call, there can be at most  $d$  calls to the  $\text{Witness}_0$  function except for the call that terminates due to  $d' = 0$  as in each of the other calls at least one index  $k \in [d]$  is removed from the tuple  $K$ . Hence,  $r_1, \dots, r_d \in [d]$  and there are at most  $d^d$  possibilities to choose these numbers. In the  $r^{\text{th}}$  call, the round number  $t_r$  is an element of  $[d']_0 \subseteq [d]_0$ , and hence, there are at most  $(d+1)^d$  possibilities to choose round numbers  $t_1, \dots, t_r$ . In each round, at most  $d$  indices  $i$  are added to the tuple  $I$ . In total, tuple  $I$  contains at most  $d^2 \cdot (d+1)$  indices, and hence, there are at most  $d^2 \cdot (d+1) \cdot n^{d^2 \cdot (d+1)}$  choices for  $I$ . Once  $I$  is fixed also the indices in  $I^* \setminus I$  are fixed because the indices added to  $I$  in Definition 26 are determined by  $I$ . The tuple  $I^*$  contains at most  $d^3 + d^2 + d$  indices. In each call  $r$ , at most  $d+1$  vectors  $x^{(r,t)}$  are generated. Hence, matrix  $A$  has at most  $d \cdot (d+1)$  columns and at most  $d^3 + d^2 + d$  rows. This yields the desired bound

$$\begin{aligned} |\mathcal{C}| &\leq d^d \cdot (d+1)^d \cdot d^2 \cdot (d+1) \cdot n^{d^2 \cdot (d+1)} \cdot 2^{d \cdot (d+1) \cdot (d^3 + d^2 + d)} \\ &\leq 2^d \cdot d^{2d} \cdot 2d^3 \cdot n^{d^2 \cdot (d+1)} \cdot 2^{d^2 \cdot (d^2 + d + 1) \cdot (d+1)} \\ &\leq 2^{(d+1)^5} \cdot d^{2d+3} \cdot n^{d^2 \cdot (d+1)}. \quad \square \end{aligned}$$

In the next step we analyze how much information of  $V$  is needed in order to perform the call  $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$ . We will see that  $V$  does not need to be revealed completely and that some randomness remains even after the necessary information to perform the call has been revealed. This is the key observation for analyzing the probability  $\mathbf{Pr}_V[E_{I^*, A, B}]$ . For this, let  $V$  be an arbitrary realization, i.e., we do not condition on the  $\text{OK}_0$ -event anymore, and fix an index  $k \in [d]$ . Only the indices  $i \in P_k$  are relevant for function  $V^k$ . Hence, we set  $I_k^* = I^* \cap P_k$

and we assume that the coefficients of  $V^k$  belonging to indices  $i \notin I_k^*$  are known. We denote this part of  $V^k$  by  $V_{I_k^*}^k$  and concentrate on the remaining part of  $V^k$  which we denote by  $V_{I_k^*}^k$ . By the construction of index  $r_k$  we know that only in the calls  $r = 1, \dots, r_k$  information about the function  $V^k$  must be available as in all subsequent calls this function is not considered anymore.

Since in the  $r^{\text{th}}$  call we consider only vectors that agree with the vectors  $x^{(r,t_r)}, \dots, x^{(r,d_r')}$  (see line 4) in the indices  $i \in I^*$ , it suffices to know all linear combinations  $V_{I_k^*}^k \cdot (a^{(r,t)}|_{I_k^*} - u|_{I_k^*})$ . However, in call  $r_k$  we would reveal too much information about  $x$  such that there would be no randomness left. Therefore, we reveal those linear combinations for all calls  $r \in [r_k - 1]$  and analyze call  $r_k$  more detailed like in the case of non-zero-preserving perturbations:

Consider call number  $r_k$  and let  $j_k \in [d_{r_k}']$  be the index such that  $k_{j_k} = k$  in round  $r_k$ , i.e.,  $V^k$  is the  $j_k^{\text{th}}$  objective function in  $K$  in round  $r_k$ . For this call the inequality  $t_{r_k} \leq j_k - 1$  holds. This is due to the fact that  $x^{(r_k, t_{r_k})}|_{P_k} = x|_{P_k}$ , i.e.,  $V^k x^{(r_k, t_{r_k})} = V^k x$ , since this is the round when  $k$  is removed from tuple  $K$ . On the other hand,  $x^{(r_k, t_{r_k})} \in \mathcal{C}_{t_{r_k}}^{(r_k)}$ , which means that  $V^{k_{t'}} x^{(r_k, t_{r_k})} < V^{k_{t'}} x$  for all indices  $t' = 1, \dots, t_{r_k}$ , where  $k_1, \dots, k_{d_{r_k}'}$  denote the indices tuple  $K$  consists of in round  $r_k$ . Hence,  $t_{r_k} < j_k$  as  $k = k_{j_k}$ .

There are only three lines where information about  $V$  is required: line 7, line 9, and line 16. For line 7 the values  $V_{I_k^*}^k \cdot (a^{(r_k, t)}|_{I_k^*} - u|_{I_k^*})$  from round  $t = d_{r_k}'$  down to round  $j_k$  are needed. In line 9 no additional information about  $V_{I_k^*}^k$  is needed since all considered vectors agree on indices  $i \in I_k^*$  with each other. For line 16 only in round  $j_k - 1$  the values  $V_{I_k^*}^k \cdot (a^{(r_k, t')}|_{I_k^*} - a^{(r_k, j_k - 1)}|_{I_k^*})$  for  $t' = t_{r_k}, \dots, j_k - 2$  are required. We write the linear combinations of  $V_{I_k^*}^k$  of the calls  $r = 1, \dots, r_k - 1$  and of call  $r_k$  into the matrices  $P_k$  and  $Q_k$  and obtain

$$P_k = \left[ a^{(1, d_1')} - u|_{I^*}, \dots, a^{(1, t_1)} - u|_{I^*}, \dots, a^{(r_k - 1, d_{r_k - 1}')} - u|_{I^*}, \dots, a^{(r_k - 1, t_{r_k - 1})} - u|_{I^*} \right] \Big|_{I_k^*} \quad \text{and}$$

$$Q_k = \left[ a^{(r_k, d_{r_k}')} - u|_{I^*}, \dots, a^{(r_k, j_k)} - u|_{I^*}, a^{(r_k, j_k - 2)} - a^{(r_k, j_k - 1)}, \dots, a^{(r_k, t_{r_k})} - a^{(r_k, j_k - 1)} \right] \Big|_{I_k^*}.$$

Using the notation  $p_k^{(r, t)} = a^{(r, t)}|_{I_k^*} - u|_{I_k^*}$  we can write both matrices as

$$P_k = \left[ p_k^{(1, d_1')}, \dots, p_k^{(1, t_1)}, \dots, p_k^{(r_k - 1, d_{r_k - 1}')} , \dots, p_k^{(r_k - 1, t_{r_k - 1})} \right] \Big|_{I_k^*} \quad \text{and}$$

$$Q_k = \left[ p_k^{(r_k, d_{r_k}')} , \dots, p_k^{(r_k, j_k)} , p_k^{(r_k, j_k - 2)} - p_k^{(r_k, j_k - 1)} , \dots, p_k^{(r_k, t_{r_k})} - p_k^{(r_k, j_k - 1)} \right] \Big|_{I_k^*}.$$

Note that the matrices  $P_k = P_k(I^*, A, u)$  and  $Q_k = Q_k(I^*, A, u)$  depend on the choice of  $u$ .

**Corollary 32.** *Let  $u \in \{0, 1\}^n$  be an arbitrary shift vector, let  $(I^*, A) \in \mathcal{C}$  be an arbitrary certificate, and let  $V$  and  $W$  be two realizations such that  $V_{I_k^*}^k = W_{I_k^*}^k$  and  $V_{I_k^*}^k \cdot q = W_{I_k^*}^k \cdot q$  for any index  $k \in [d]$  and any column  $q$  of one of the matrices  $P_k(I^*, A, u)$  and  $Q_k(I^*, A, u)$ . Then, for any  $\varepsilon$ -box  $B \in \mathbb{B}_\varepsilon$  the calls  $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$  and  $\text{Witness}_0(W, [d], 1, I^*, A, \mathcal{S}, B, u)$  return the same result.*

In the remainder of this section we assume that  $V_{I_k^*}^k$  and the  $\varepsilon$ -box  $B$  are fixed. In accordance with Corollary 32, the output of the call  $\text{Witness}_0(V, [d], 1, I^*, A, \mathcal{S}, B, u)$  is determined if the linear combinations of  $V_{I_k^*}^k$  given by the columns of the matrices  $P_k$  and  $Q_k$  are known in addition. We are interested in the event  $E_{I^*, A, B}$ , i.e., in the event that the output is a set  $\{x'\}$  such that  $V^{1..d} \cdot (x' - u^*(I^*, A)) \in B$ . Since  $(I^*, A)$  is a  $V$ -certificate of  $x$  for some  $V$  and  $x$ , the output is always of the form  $\{x'\}$  due to Lemma 28. Hence, event  $E_{I^*, A, B}$  occurs if and only if for all indices  $k$  the relation  $V_{I_k^*}^k \cdot (x' - u^*(I^*, A))|_{I_k^*} \in C_k$  holds for some interval  $C_k$  of length  $\varepsilon$  that depends on the linear combinations of  $V_{I_\ell^*}^k$  given by  $P_\ell$  and  $Q_\ell$  for all indices  $\ell \in [d]$ .

**Lemma 33.** For any fixed index  $k \in [d]$  the columns of matrix  $P_k(I^*, A, u^*(I^*, A))$ , of matrix  $Q_k(I^*, A, u^*(I^*, A))$ , and the vector  $p_k^{(r_k, t_{r_k})}$  are linearly independent.

*Proof.* Consider the square matrix  $\hat{Q}_k$  consisting of the vectors  $p_k^{(r, t)}$ , for  $r \in [r_k]$  and  $t \in \{t_r, \dots, d'_r\}$ . Due to Lemma 27 and due to the construction of  $u^*(I^*, A)$  (see Equation 3) matrix  $\hat{Q}_k$  is a lower triangular matrix and the elements of the principal diagonal are from the set  $\{-1, 1\}$ . Hence,  $|\det \hat{Q}_k| = 1$ , i.e., the vectors  $p_k^{(r, t)}$  are linearly independent.

The columns of matrix  $P_k$  are pairwise distinct columns of matrix  $\hat{Q}_k$ . The columns of matrix  $Q_k$  and vector  $p_k^{(r_k, t_{r_k})}$  are linear combinations of the vectors  $p_k^{(r_k, t_{r_k})}, \dots, p_k^{(r_k, d'_{r_k})}$ . As the vectors  $p_k^{(r, t)}$  are linearly independent, it suffices to show that the columns of matrix  $Q_k$  and vector  $p_k^{(r_k, t_{r_k})}$  are linearly independent. For this consider an arbitrary linear combination of the columns of matrix  $Q_k$  and vector  $p_k^{(r_k, t_{r_k})}$  that computes to zero:

$$\sum_{t=j_k}^{d'_{r_k}} \lambda_k^{(t)} \cdot p_k^{(r_k, t)} + \sum_{t=t_{r_k}}^{j_k-2} \lambda_k^{(t)} \cdot (p_k^{(r_k, t)} - p_k^{(r_k, j_k-1)}) + \mu_k \cdot p_k^{(r_k, t_{r_k})} = 0.$$

If  $t_{r_k} = j_k - 1$ , then this equation is equivalent to

$$\sum_{t=t_{r_k}+1}^{d'_{r_k}} \lambda_k^{(t)} \cdot p_k^{(r_k, t)} + \mu_k \cdot p_k^{(r_k, t_{r_k})} = 0,$$

i.e., all coefficients are zero due to the linear independence of the vectors  $p_k^{(r_k, t)}$ . If  $t_{r_k} < j_k - 1$ , then the equation is equivalent to

$$\sum_{t=j_k}^{d'_{r_k}} \lambda_k^{(t)} \cdot p_k^{(r_k, t)} + \sum_{t=t_{r_k}+1}^{j_k-2} \lambda_k^{(t)} \cdot p_k^{(r_k, t)} - \left( \sum_{t=t_{r_k}}^{j_k-2} \lambda_k^{(t)} \right) \cdot p_k^{(r_k, j_k-1)} + (\lambda_k^{(t_{r_k})} + \mu_k) \cdot p_k^{(r_k, t_{r_k})} = 0.$$

Due to the linear independence of the vectors  $p_k^{(r_k, t)}$  this implies  $\lambda_k^{(t)} = 0$  for  $t \in \{t_{r_k} + 1, \dots, j_k - 2\} \cup \{j_k, \dots, d'_{r_k}\}$ ,  $\sum_{t=t_{r_k}}^{j_k-2} \lambda_k^{(t)} = 0$ , and  $\lambda_k^{(t_{r_k})} + \mu_k = 0$  and we obtain that all coefficients are zero. In both cases, the linear independence of the columns of  $Q_k$  and the vector  $p_k^{(r_k, t_{r_k})}$  follows.  $\square$

**Corollary 34.** Let  $(I^*, A) \in \mathcal{C}$  be an arbitrary certificate and let  $B \in \mathbb{B}_\varepsilon$  be an arbitrary  $\varepsilon$ -box. Then,

$$\Pr_V[E_{I^*, A, B}] \leq (2\gamma)^{\gamma-d} \cdot \phi^\gamma \cdot \varepsilon^d,$$

and

$$\Pr_V[E_{I^*, A, B}] \leq 2^d \cdot \gamma^{\gamma-d} \cdot \phi^d \cdot \varepsilon^d$$

in case of quasiconcave densities, where  $\gamma = d^3 + d^2 + d$ .

*Proof.* For fixed index  $k \in [d]$  we write the columns of matrix  $P_k$ , of matrix  $Q_k$ , and vector  $p_k^{(r_k, t_{r_k})}$  into one matrix  $Q_k$  and consider the matrix

$$Q' = \begin{bmatrix} Q_1 & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbb{O} \\ \mathbb{O} & \dots & \mathbb{O} & Q_d \end{bmatrix} \in \{-1, 0, 1\}^{|I^*| \times \ell}$$

for some integer  $\ell \leq |I^*|$ . By Lemma 33 this matrix has (full) rank  $\ell$ . Now, permute the columns of  $Q'$  to obtain a matrix  $Q$  where the last  $d$  columns belong to the last column of one of the matrices  $Q_k$ . For any index  $i \in I^*$  let  $X_i = V_i^k$  be the  $i^{\text{th}}$  coefficient of  $V^k$  where  $k$  is the unique index such that  $i \in P_k$ . Event  $E_{I^*, A, B}$  holds if and only if the  $d$  linear combinations of the variables  $X_i$  given by the last  $d$  columns of  $Q$  fall into a  $d$ -dimensional hypercube  $C$  depending on the linear combinations of the variables  $X_i$  given by the remaining columns. The claim follows by applying Theorem 35 for matrix  $A = Q^T$  and due to the fact that  $|I^*| \leq \gamma$  (see proof of Lemma 31).  $\square$

*Proof of Theorem 1.* We begin the proof by showing that the  $\text{OK}_0$ -event is likely to happen. For any index  $t \in [d]$  and any solutions  $x, y \in \mathcal{S}$  such that  $x|_{P_t} \neq y|_{P_t}$  the probability that  $|V^t x - V^t y| \leq \varepsilon$  is bounded by  $2\phi\varepsilon$ . To see this, choose one index  $i \in P_t$  such that  $x_i \neq y_i$  and fix all coefficients  $V_j^t$  for  $j \neq i$ . Then, the value  $V_i^t$  must fall into an interval of length  $2\varepsilon$ . A union bound over all indices  $t \in [d]$  and over all pairs  $(x, y) \in \mathcal{S} \times \mathcal{S}$  such that  $x|_{P_t} \neq y|_{P_t}$  yields  $\Pr_V[\overline{\text{OK}_0(V)}] \leq 2^{2n+1}d\phi\varepsilon$ . For  $\gamma = d^3 + d^2 + d$ , we set  $s = (2\gamma)^{\gamma-d} \cdot \phi^\gamma$  for general density functions and  $s = 2^d \cdot \gamma^{\gamma-d} \cdot \phi^d$  in the case of quasiconcave density functions. Then, we obtain

$$\begin{aligned} \mathbf{E}_V[\text{PO}(V)] &\leq \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} \Pr_V[E_{I^*, A, B}] + 2^n \cdot \Pr_V[\overline{\text{OK}_0(V)}] \\ &\leq \sum_{(I^*, A) \in \mathcal{C}} \sum_{B \in \mathbb{B}_\varepsilon} s \cdot \varepsilon^d + 2^n \cdot 2^{2n+1}d\phi\varepsilon \\ &\leq 2^{(d+1)^5} \cdot d^{2d+3} \cdot n^{d^2 \cdot (d+1)} \cdot \left(\frac{2n}{\varepsilon}\right)^d \cdot s \cdot \varepsilon^d + 2^{3n+1}d\phi\varepsilon \\ &= 2^{(d+1)^5+d} \cdot d^{2d+3} \cdot n^{d^3+d^2+d} \cdot s + 2^{3n+1}d\phi\varepsilon. \end{aligned}$$

The first inequality is due to Corollary 30. The second inequality is due to Corollary 34. The third inequality stems from Lemma 31. Since this bound is true for arbitrarily small  $\varepsilon > 0$ , the correctness of Theorem 1 follows.  $\square$

## 6 Some Probability Theory

**Theorem 35.** *Let  $X_1, \dots, X_n$  be independent random variables, each with a probability density function  $f_i: [-1, 1] \rightarrow [0, \phi]$ . Let  $m \in [n]$ , let  $A \in \{-1, 0, 1\}^{m \times n}$  be a matrix of full rank, let  $k \in [m-1]$  be an integer, let  $(Y_1, \dots, Y_{m-k}, Z_1, \dots, Z_k)^T = A \cdot (X_1, \dots, X_n)^T$  be the linear combinations of  $X_1, \dots, X_n$  given by  $A$ , and let  $C$  be a function mapping a tuple  $(y_1, \dots, y_{m-k}) \in \mathbb{R}^{m-k}$  to a hypercube  $C(y_1, \dots, y_{m-k}) \subseteq \mathbb{R}^k$  with side length  $\varepsilon$ . Then,*

$$\Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k})] \leq (2n)^{n-k} \cdot \phi^n \cdot \varepsilon^k.$$

*If all densities  $f_i$  are quasiconcave, then even the stronger bound*

$$\Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k})] \leq 2^k \cdot n^{n-k} \cdot \phi^k \cdot \varepsilon^k$$

*holds.*

*Proof.* Without loss of generality let  $m = n$ , i.e., matrix  $A$  is a square matrix where  $\det A \neq 0$ . Otherwise, we could insert  $m - n$  linearly independent rows from  $\{0, 1\}^n$  after row  $m - k$  until matrix  $A$  is of that form. That way we get new random variables  $Y_{m-k+1}, \dots, Y_{n-k}$ . Then, we can extend the domain of  $C$  to  $\mathbb{R}^{n-k}$  by the definition  $C(y_1, \dots, y_{n-k}) := C(y_1, \dots, y_{m-k})$ , which shows that it is sufficient to deal with the case  $m = n$  because  $\Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{m-k})] = \Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{n-k})]$ .



As matrix  $A$  is a full-rank square matrix, its inverse  $A^{-1}$  exists and we can write

$$\begin{aligned}
\Pr[(Z_1, \dots, Z_k) \in C(Y_1, \dots, Y_{n-k})] &= \int_{y \in \mathbb{R}^{n-k}} \int_{z \in C(y)} f_{Y,Z}(y, z) \, dz \, dy \\
&= \int_{y \in \mathbb{R}^{n-k}} \int_{z \in C(y)} |\det(A^{-1})| \cdot f_X(A^{-1} \cdot (y, z)) \, dz \, dy \\
&\leq \int_{y \in \mathbb{R}^{n-k}} \int_{z \in C(y)} f_X(A^{-1} \cdot (y, z)) \, dz \, dy \\
&\leq \varepsilon^k \cdot \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) \, dy,
\end{aligned}$$

where  $f_{Y,Z}$  denotes the common density of the variables  $Y_1, \dots, Y_{n-k}, Z_1, \dots, Z_k$  and  $f_X = \prod_{i=1}^n f_i$  denotes the common density of the variables  $X_1, \dots, X_n$ . The second equality is due to a change of variables, the first inequality stems from the fact that  $|\det(A^{-1})| = |1/\det A| \in (0, 1)$  since  $A$  is an integer matrix.

In general, we can bound the integral in the formula above by

$$\begin{aligned}
\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) \, dy &\leq \int_{y \in [-n, n]^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) \, dy \\
&\leq \int_{y \in [-n, n]^{n-k}} \phi^n \, dy = (2n)^{n-k} \cdot \phi^n,
\end{aligned}$$

where the first inequality is due to the fact that all variables  $Y_i$  can only take values in the interval  $[-n, n]$  as all entries of matrix  $A$  are from  $\{-1, 0, 1\}$ .

To prove the statement about quasiconcave functions we first consider arbitrary rectangular functions, i.e., functions that are constant on a given interval, and zero otherwise. This will be the main part of our analysis. Afterwards, we analyze sums of rectangular functions and, finally, we show that quasiconcave functions can be approximated by such sums.

**Lemma 36.** *For  $i \in [n]$  let  $\phi_i \geq 0$ , let  $I_i \subseteq \mathbb{R}$  be an interval of length  $\ell_i$ , and let  $f_i: \mathbb{R} \rightarrow \mathbb{R}$  be the function*

$$f_i(x) = \begin{cases} \phi_i & : x \in I_i, \\ 0 & : \text{otherwise.} \end{cases}$$

*Moreover, let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be the function  $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$  and let  $A \in \{-1, 0, 1\}^{n \times n}$  be an invertible matrix. Then,*

$$\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) \, dy \leq 2^k \cdot (n-k)! \cdot \chi \cdot \sum_I \prod_{i \notin I} \ell_i$$

*where  $\chi = \prod_{i=1}^n \phi_i$  and where the sum runs over all tuples  $I = (i_1, \dots, i_k)$  where  $1 \leq i_1 < \dots < i_k \leq n$ .*

*Proof.* Function  $f$  takes the value  $\chi$  on the  $n$ -dimensional hyper-cuboid  $Q = \prod_{i=1}^n I_i$  and is zero otherwise. Hence,

$$\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) \, dy = \chi \cdot \text{vol}(Q')$$

for

$$\begin{aligned}
Q' &= \{y \in \mathbb{R}^{n-k} : \exists z \in \mathbb{R}^k \text{ such that } A^{-1} \cdot (y, z) \in Q\} \\
&= \{y \in \mathbb{R}^{n-k} : \exists z \in \mathbb{R}^k \exists x \in Q \text{ such that } (y, z) = A \cdot x\}
\end{aligned}$$

$$= (P \cdot A)(Q),$$

where  $P = [\mathbb{I}_{n-k}, \mathbb{O}_{(n-k) \times k}]$  is the projection matrix that removes the last  $k$  entries from a vector of length  $n$ . In the remainder of this proof we bound the volume of  $M(Q)$  where  $M = P \cdot A \in \{-1, 0, 1\}^{(n-k) \times n}$ . Let  $a_i =: c_i^0$  and  $b_i =: c_i^1$  be the left and the right bound of interval  $I_i$ , respectively. For an index tuple  $I = (i_1, \dots, i_k)$ ,  $1 \leq i_1 < \dots < i_k \leq n$ , and a bit tuple  $J = (j_1, \dots, j_k) \in \{0, 1\}^k$ , let

$$F_I^J = \prod_{i=1}^n \begin{cases} \{c_{i_t}^{j_t}\} & : i = i_t, \\ I_i & : i \notin I, \end{cases}$$

be one of the  $2^k \cdot \binom{n}{k}$   $(n-k)$ -dimensional faces of  $Q$ . We show that  $M(Q) \subseteq \bigcup_I \bigcup_J M(F_I^J)$ . Let  $y \in M(Q)$ , i.e., there is a vector  $x \in Q$  such that  $y = M \cdot x$ . Now, consider the polytope

$$R = \{(x', s') \in \mathbb{R}^n \times \mathbb{R}^n : M \cdot x' = y', x' + s' = b', \text{ and } x', s' \geq 0\},$$

where  $y' = y - M \cdot a$  and  $b' = b - a$  for  $a = (a_1, \dots, a_n)$  and  $b = (b_1, \dots, b_n)$ . This polytope is bounded and non-empty since  $(x', s') \in R$  for  $x' = x - a$  and  $s' = b - x$ . Consequently, there exists a basic feasible solution  $(x^*, s^*)$ . As there are  $2n$  variables and  $2n - k$  constraints, this solution has at least  $k$  zero-entries, i.e., there are indices  $1 \leq i_1 < \dots < i_k \leq n$  such that either  $x_{i_t}^* = 0$  (in that case set  $j_t = 0$ ) or  $x_{i_t}^* = b_{i_t}'$  (in that case set  $j_t = 1$ ) for any  $t \in [k]$ . Now, consider the vector  $\hat{x} = x^* + a \in [0, b'] + a = Q$ . We obtain  $M \cdot \hat{x} = y$  and  $\hat{x}_{i_t} = c_{i_t}^{j_t}$  for all  $t \in [k]$ . Hence,  $x \in F_I^J$  for  $I = (i_1, \dots, i_k)$  and  $J = (j_1, \dots, j_k)$ , and thus  $y \in M(F_I^J)$ .

Due to this observation we can bound the volume of  $M(Q)$  by  $\sum_I \sum_J \text{vol}(M(F_I^J))$ . It remains to show how to bound the volume  $\text{vol}(M(F_I^J))$ . For the sake of simplicity we only consider  $I = (n-k+1, \dots, n)$  in the following analysis. Let  $\phi^J : \mathbb{R}^{n-k} \rightarrow F_I^J$  be the function  $\phi^J(x) = T \cdot x + v^J$ , where  $T = [\mathbb{I}_{n-k}, \mathbb{O}_{(n-k) \times k}]^T$  and  $v^J = (0, \dots, 0, c_{n-k+1}^{j_1}, \dots, c_n^{j_k})$ . Using function  $\phi^J$  is the canonical way to describe the affine subspace defined by face  $F_I^J$ : it adds the fixed coordinates of  $F_I^J$  to a given vector of length  $n-k$ . Hence, function  $\phi^J$ , restricted to the domain  $F' = \prod_{i=1}^{n-k} I_i$ , is bijective. With  $\psi = M \circ \phi^J$  we obtain

$$\begin{aligned} \text{vol}(M(F_I^J)) &= \int_{\psi(F')} 1 \, dx = \int_{F'} |\det D\psi(x)| \, dx = \int_{F'} |\det(M \cdot T)| \, dx \\ &= |\det(M \cdot T)| \cdot \text{vol}(F') = |\det(M \cdot T)| \cdot \prod_{i=1}^{n-k} \ell_i. \end{aligned}$$

Matrix  $M \cdot T = P \cdot A \cdot T$  is a  $(n-k) \times (n-k)$ -submatrix of  $A$ . Thus,  $|\det(M \cdot T)| \leq (n-k)!$ , and we obtain the bound

$$\begin{aligned} \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) \, dy &= \chi \cdot \text{vol}(M(Q)) \leq \chi \cdot \sum_I \sum_J \text{vol}(M(F_I^J)) \\ &\leq \chi \cdot \sum_I \sum_J (n-k)! \cdot \prod_{i \notin I} \ell_i \\ &= 2^k \cdot (n-k)! \cdot \chi \cdot \sum_I \prod_{i \notin I} \ell_i. \end{aligned} \quad \square$$

In the next step we generalize the statement of Lemma 36 to sums of rectangular functions.

**Corollary 37.** *Let  $N_1, \dots, N_n$  be positive integers, let  $\phi_{i,k} \geq 0$  be a non-negative real, let  $I_{i,k} \subseteq \mathbb{R}$  be an interval of length  $\ell_{i,k}$ , and let  $f_{i,k} : \mathbb{R} \rightarrow \mathbb{R}$  be the function*

$$f_{i,k}(x) = \begin{cases} \phi_{i,k} & : x \in I_{i,k}, \\ 0 & : \text{otherwise}, \end{cases}$$

$i \in [n]$ ,  $k \in [N_i]$ . Furthermore, let  $f_i: \mathbb{R} \rightarrow \mathbb{R}$  be the function  $f_i = \sum_{k=1}^{N_i} f_{i,k}$ , let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be the function  $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$ , and let  $A \in \{-1, 0, 1\}^{n \times n}$  be an invertible matrix. Then,

$$\int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) dy \leq 2^k \cdot (n-k)! \cdot \sum_I \left( \prod_{i \notin I} \sigma_i \right) \cdot \left( \prod_{i \in I} \chi_i \right)$$

where  $\sigma_i = \sum_{k=1}^{N_i} \phi_{i,k} \cdot \ell_{i,k}$  and  $\chi_i = \sum_{k=1}^{N_i} \phi_{i,k}$  and where the first sum runs over all tuples  $I = (i_1, \dots, i_k)$  for which  $1 \leq i_1 < \dots < i_k \leq n$ .

*Proof.* For indices  $k_i \in [N_i]$  let  $f_{k_1, \dots, k_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{i, k_i}(x_i)$ . This function is of the form assumed in Lemma 36 and takes only values zero and  $\chi_{k_1, \dots, k_n} = \prod_{i=1}^n \phi_{i, k_i}$ . We can write function  $f$  as

$$\begin{aligned} f(x_1, \dots, x_n) &= \prod_{i=1}^n f_i(x_i) = \prod_{i=1}^n \sum_{k_i=1}^{N_i} f_{i, k_i}(x_i) = \sum_{k_1=1}^{N_1} \dots \sum_{k_n=1}^{N_n} \prod_{i=1}^n f_{i, k_i}(x_i) \\ &= \sum_{k_1=1}^{N_1} \dots \sum_{k_n=1}^{N_n} f_{k_1, \dots, k_n}(x_1, \dots, x_n). \end{aligned}$$

For the sake of simplicity we sometimes write  $(\sum_{k_j=1}^{N_j})_{j \in (j_1, \dots, j_\ell)}$  instead of  $\sum_{k_{j_1}=1}^{N_{j_1}} \dots \sum_{k_{j_\ell}=1}^{N_{j_\ell}}$  and we drop the bounds of the sums if they are clear. Now, we can bound the integral as follows:

$$\begin{aligned} \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f(A^{-1} \cdot (y, z)) dy &= \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} \sum_{k_1=1}^{N_1} \dots \sum_{k_n=1}^{N_n} f_{k_1, \dots, k_n}(A^{-1} \cdot (y, z)) dy \\ &\leq \sum_{k_1=1}^{N_1} \dots \sum_{k_n=1}^{N_n} \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_{k_1, \dots, k_n}(A^{-1} \cdot (y, z)) dy \\ &\leq \sum_{k_1=1}^{N_1} \dots \sum_{k_n=1}^{N_n} \left( 2^k \cdot (n-k)! \cdot \chi_{k_1, \dots, k_n} \cdot \sum_I \prod_{i \notin I} \ell_{i, k_i} \right) \\ &= 2^k \cdot (n-k)! \cdot \sum_I \left( \sum_{k_i} \right)_{i \in I} \left( \prod_{i \in I} \phi_{i, k_i} \cdot \left( \sum_{k_i} \right)_{i \notin I} \prod_{i \notin I} \phi_{i, k_i} \cdot \ell_{i, k_i} \right) \\ &= 2^k \cdot (n-k)! \cdot \sum_I \left( \sum_{k_i} \right)_{i \in I} \left( \prod_{i \in I} \phi_{i, k_i} \cdot \prod_{i \notin I} \sum_{k_i} \phi_{i, k_i} \cdot \ell_{i, k_i} \right) \\ &= 2^k \cdot (n-k)! \cdot \sum_I \left( \prod_{i \notin I} \sum_{k_i} \phi_{i, k_i} \cdot \ell_{i, k_i} \cdot \left( \sum_{k_i} \right)_{i \in I} \prod_{i \in I} \phi_{i, k_i} \right) \\ &= 2^k \cdot (n-k)! \cdot \sum_I \left( \prod_{i \notin I} \sum_{k_i} \phi_{i, k_i} \cdot \ell_{i, k_i} \cdot \prod_{i \in I} \sum_{k_i} \phi_{i, k_i} \right) \\ &= 2^k \cdot (n-k)! \cdot \sum_I \left( \prod_{i \notin I} \sigma_i \right) \cdot \left( \prod_{i \in I} \chi_i \right), \end{aligned}$$

where the second inequality is due to Lemma 36.  $\square$

To finish the proof of Theorem 35 we round the probability densities  $f_i$  as follows: For an arbitrarily small positive real  $\delta$  let  $g_i := \lceil f_i / \delta \rceil \cdot \delta$ , i.e., we round  $f_i$  up to the next integral multiple

of  $\delta$ . As the densities  $f_i$  are quasiconcave, there is a decomposition of  $g_i$  such that  $g_i = \sum_{k=1}^{N_i} f_{i,k}$  where

$$f_{i,k} = \begin{cases} \phi_{i,k} & : x \in I_{i,k}, \\ 0 & : \text{otherwise}, \end{cases} \quad \text{and} \quad \chi_i := \sum_{k=1}^{N_i} \phi_{i,k} = \max_{x \in [-1,1]} g_i(x),$$

where  $I_{i,k}$  are intervals of length  $\ell_{i,k}$  and  $\phi_{i,k}$  are positive reals. The second property is the interesting one and stems from the quasiconcaveness of  $f_i$ . Informally speaking the two-dimensional shape bounded by the horizontal axis and the graph of  $g_i$  is a stack of rectangles aligned with axes. Therefore, the sum  $\chi_i$  of the rectangles' heights which appears in the formula of Corollary 37 is approximately  $\phi$ . Without the quasiconcaveness  $\chi_i$  might be unbounded.

Applying Corollary 37, we obtain

$$\begin{aligned} \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) dy &= \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} \prod_{i=1}^n f_i((A^{-1} \cdot (y, z))_i) dy \\ &\leq \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} \prod_{i=1}^n g_i((A^{-1} \cdot (y, z))_i) dy \\ &\leq 2^k \cdot (n-k)! \cdot \sum_I \left( \prod_{i \notin I} \sum_{k_i=1}^{N_i} \phi_{i,k_i} \cdot \ell_{i,k_i} \right) \cdot \left( \prod_{i \in I} \chi_i \right) \\ &= 2^k \cdot (n-k)! \cdot \sum_I \left( \prod_{i \notin I} \int_{[-1,1]} g_i dx \right) \cdot \left( \prod_{i \in I} \chi_i \right). \end{aligned}$$

Since  $0 \leq \int_{[-1,1]} g_i dx \leq \int_{[-1,1]} (f_i + \delta) dx = 1 + 2\delta$  and  $0 \leq \chi_i \leq \sup_{x \in [-1,1]} f_i(x) + \delta \leq \phi + \delta$ , this implies

$$\begin{aligned} \int_{y \in \mathbb{R}^{n-k}} \max_{z \in \mathbb{R}^k} f_X(A^{-1} \cdot (y, z)) dy &\leq 2^k \cdot (n-k)! \cdot \sum_I \left( \prod_{i \notin I} (1 + 2\delta) \right) \cdot \left( \prod_{i \in I} (\phi + \delta) \right) \\ &= 2^k \cdot (n-k)! \cdot \sum_I (1 + 2\delta)^{n-k} \cdot (\phi + \delta)^k \\ &= 2^k \cdot (n-k)! \cdot \binom{n}{k} \cdot (1 + 2\delta)^{n-k} \cdot (\phi + \delta)^k \\ &\leq 2^k \cdot n^{n-k} \cdot (1 + 2\delta)^{n-k} \cdot (\phi + \delta)^k. \end{aligned}$$

As this bound is true for arbitrarily small reals  $\delta > 0$ , we obtain the desired bound.  $\square$

## 7 Conclusions and Open Problems

With the techniques developed in this paper we settled two questions posed by Moitra and O'Donnell [13]: For quasiconcave densities we showed that the exponent of  $\phi$  in the bound for the smoothed number of Pareto-optimal solutions is exactly  $d$ . Moreover, we significantly improved on the previously best known bound for higher moments of the smoothed number of Pareto-optima by Röglin and Teng [17].

Maybe even more interesting are our results for the model of zero-preserving perturbations suggested by Spielman and Teng [20] and Beier and Vöcking [4]. For this model we proved the first non-trivial bound on the smoothed number of Pareto-optimal solutions. We showed that this result can be used to analyze multiobjective optimization problems with polynomial and even more general objective functions. Furthermore, our result implies that the smoothed running time of the algorithm proposed by Berger et al. [5] to compute a path trade in a routing network is

polynomially bounded for any constant number of autonomous systems. We believe that there are many more such applications of our result in the area of multi-objective optimization.

There are several interesting open questions. First of all it would be interesting to find asymptotically tight bounds for the smoothed number of Pareto-optimal solutions. There is still a gap between our upper bound of  $O(n^{2d}\phi^d)$  for quasiconcave  $\phi$ -smooth instances and the best lower bound of  $\Omega(n^{d-1.5}\phi^d)$  [10]. Only for the case  $d = 1$ , we could show that the upper bound is tight [6].

Especially for zero-preserving perturbations there is still a lot of work to do. We conjecture that our techniques can be extended to also bound higher moments of the smoothed number of Pareto-optima for  $\phi$ -smooth instances with zero-preserving perturbations. However, we feel that even our bound for the first moment is too pessimistic as we do not have a lower bound showing that setting coefficients to zero can lead to larger Pareto-sets. It would be very interesting to either prove a lower bound that shows that zero-preserving perturbations can lead to larger Pareto-sets than non-zero-preserving perturbations or to prove a better upper bound for zero-preserving perturbations.

It would also be interesting to know whether our results extend to sets  $\mathcal{S} \subseteq \{-m, \dots, m\}^n$  of solutions for a constant  $m > 1$ . We conjecture that our results still hold with worse constants that depend on  $m$ .

## References

- [1] René Beier. *Probabilistic Analysis of Discrete Optimization Problems*. PhD thesis, Universität des Saarlandes, 2004.
- [2] René Beier, Heiko Röglin, and Berthold Vöcking. The smoothed number of Pareto optimal solutions in bicriteria integer optimization. In *Proceedings of the 12th International Conference on Integer Programming and Combinatorial Optimization (IPCO)*, pages 53–67, 2007.
- [3] René Beier and Berthold Vöcking. Random knapsack in expected polynomial time. *Journal of Computer and System Sciences*, 69(3):306–329, 2004.
- [4] René Beier and Berthold Vöcking. Typical properties of winners and losers in discrete optimization. *SIAM Journal on Computing*, 35(4):855–881, 2006.
- [5] André Berger, Heiko Röglin, and Ruben van der Zwaan. Path trading: Fast algorithms, smoothed analysis, and hardness results. In *Proceedings of the 10th International Symposium on Experimental Algorithms (SEA)*, pages 43–53, 2011.
- [6] Tobias Brunsch and Heiko Röglin. Lower bounds for the smoothed number of pareto optimal solutions. In *Proceedings of the 8th Annual Conference on Theory and Applications of Models of Computation (TAMC)*, pages 416–427, 2011.
- [7] H. William Corley and I. Douglas Moon. Shortest paths in networks with vector weights. *Journal of Optimization Theory and Application*, 46(1):79–86, 1985.
- [8] Matthias Ehrgott. Integer solutions of multicriteria network flow problems. *Investigacao Operacional*, 19:229–243, 1999.
- [9] Matthias Ehrgott. *Multicriteria Optimization*. Springer, 2005.
- [10] Navin Goyal and Luis Rademacher. Lower bounds for the average and smoothed number of pareto optima. *CoRR*, abs/1107.3876, 2011.

- [11] Pierre Hansen. Bicriterion path problems. In *Multiple Criteria Decision Making: Theory and Applications*, volume 177 of *Lecture Notes in Economics and Mathematical Systems*, pages 109–127, 1980.
- [12] Kathrin Klamroth and Margaret M. Wiecek. Dynamic programming approaches to the multiple criteria knapsack problem. *Naval Research Logistics*, 47(1):57–76, 2000.
- [13] Ankur Moitra and Ryan O’Donnell. Pareto optimal solutions for smoothed analysts. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 225–234, 2011.
- [14] Matthias Müller-Hannemann and Karsten Weihe. Pareto shortest paths is often feasible in practice. In *Proceedings of the 5th International Workshop on Algorithm Engineering (WAE)*, pages 185–198, 2001.
- [15] Adli Mustafa and Mark Goh. Finding integer efficient solutions for bicriteria and tricriteria network flow problems using dinas. *Computers & Operations Research*, 25(2):139–157, 1998.
- [16] George L. Nemhauser and Zev Ullmann. Discrete dynamic programming and capital allocation. *Management Science*, 15(9):494–505, 1969.
- [17] Heiko Röglin and Shang-Hua Teng. Smoothed analysis of multiobjective optimization. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 681–690, 2009.
- [18] Arvind Sankar, Daniel A. Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM Journal on Matrix Analysis Applications*, 28(2):446–476, 2006.
- [19] Anders J. V. Skriver and Kim Allan Andersen. A label correcting approach for solving bicriterion shortest-path problems. *Computers & Operations Research*, 27(6):507–524, 2000.
- [20] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004.
- [21] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis: an attempt to explain the behavior of algorithms in practice. *Communications of the ACM*, 52(10):76–84, 2009.